

**Kieli, kääntäminen ja instituutio – onko
institutionaalisten tekstien kieli yksinkertaisempaa
käännöksissä?**

Elina Ojala

Tampereen yliopisto

Kieli-, käännös- ja kirjallisuustieteiden yksikkö

Monikielisen viestinnän ja käännöstieteen koulutusohjelma

Työkielet suomi-englanti

Pro gradu -tutkielma

Toukokuu 2015

Tampereen yliopisto

Monikielisen viestinnän ja käännöstieteen koulutusohjelma, työkielet suomi-englanti

Kieli-, käännös- ja kirjallisuustieteiden yksikkö

Ojala, Elina: Kieli, kääntäminen ja instituutio – onko institutionaalisten tekstien kieli yksinkertaisempaa käännöksissä?

Pro gradu -tutkielma, 61 sivua, liitteet 5 sivua, englanninkielinen lyhennelmä 13 sivua.
Toukokuu 2015

Tutkimus käsittelee institutionaalisia käännöksiä käännösuniversaalihypoteesiteorian näkökulmasta. Tutkimuksessa sovelletaan korpusmenetelmiä lain ja hallinnon tekstien tutkimuksessa. Tutkimuksen tavoitteena on selvittää, päteekö käännösuniversaaliksi ehdotettu yksinkertaistumishypoteesi myös lain ja hallinnon piiriin kuuluvissa institutionaalisten tekstien käännöksissä.

Tutkielman aineisto koostuu Euroopan unionin suomenkielisistä asetuksista sekä suomalaisista säädöksistä, joista on koottu komparatiiviset korpuksat ja joille on tehty sanaluokkajäsennys. Näistä teksteistä muodostettuja verrannollisia korpuksia analysoimalla tutkitaan yksinkertaistumishypoteesia korpuksien sana-sane -suhdetta sekä leksikaalista tiheyttä vertailemalla. Tämän lisäksi korpuksien yleisiä kielellisiä piirteitä, kuten virkkeiden ja lauseiden pituuksia, vertaillaan. Aikaisempaan tutkimukseen perustuen hypoteesina on, että suomennosten sana-sane -suhde on matalampi, ja että niiden leksikaalinen tiheys alhaisempi. Toisin sanoen oletetaan, että merkkejä käännöskielen yksinkertaistumista on suomenkielisissä teksteissä havaittavissa.

Korpuksien sana-sane -suhdetta ja leksikaalista tiheyttä vertailemalla ei tutkimuksessa löydetty viitteitä käännöskielen yksinkertaistumisesta. Tutkimuksen tulokset eivät siis tue yksinkertaistumishypoteesia. Tutkimuksen tulokset ovat näin ollen ristiriidassa aikaisemman käännösuniversaalihypoteeseja koskevan tutkimuksen kanssa. On kuitenkin otettava huomioon, että aikaisempaa tutkimusta ei ole tehty lain ja hallinnon teksteistä koostuvilla korpuksilla.

Yksi selitys saaduille tutkimustuloksille on käännösuniversaalihypoteesien mahdollinen tekstilajikohtaisuus sekä se, että lain ja hallinnon teksteillä on keskenään hyvin samanlaisia piirteitä kielestä riippumatta. Tekstilajikohtaisista tutkimuksista sekä konkreettisten käännösnormien, käännösmuistien sekä tietokoneavusteisten käännöstyökalujen vaikutuksesta käännösuniversaalien esiintymiseen olisikin mielekästä tutkia tulevaisuudessa.

Avainsanat: *Euroopan unioni, institutionaalinen kääntäminen, käännösuniversaaliteoria, korpus tutkimus, lakitekstit, sana-sane -suhde, Suomen lainsäädäntö, yksinkertaistumishypoteesi*

SISÄLLYSLUETTELO

1. JOHDANTO.....	1
1.1 Tutkimuksen aihe ja tutkimuskysymys.....	1
1.2 Tutkielman rakenne	2
2. INSTITUUTIO, SEN TEKSTIT JA NIIDEN KÄÄNTÄMINEN.....	4
2.1 Instituutio käännöstieteen näkökulmasta	4
2.2 Instituutionaalisten tekstien erityispiirteet ja kääntäminen	5
2.3 Suomen lainsäädäntö ja lainvalmisteluprosessi	8
2.3.1 Lainsäädännön prosessit.....	9
2.3.2 Kielelliset standardit ja kääntäminen osana Suomen lainsäädäntöprosessia	11
2.4 Euroopan unionin lainsäädäntö.....	12
2.4.1 Euroopan unionin lainsäädäntöprosessi	12
2.4.2 Monikielinen lainsäädäntö ja kääntäminen osana lainsäädäntöprosessia Euroopan unionissa	13
2.5 Yhteenveto: instituutiot ja niiden tekstit.....	15
3. KÄÄNNÖSUNIVERSAALITEORIA JA KÄÄNNÖSSUOMEN TUTKIMUS.....	18
3.1 Käännösuniversaalliteoria.....	18
3.2 Käännösuniversaalliteorian kritiikki	21
3.3 Käännössuomen aikaisempaa tutkimusta	22
4. KORPUSLINGVISTIIKKA.....	25
4.1 Korpuksen määritelmä ja korpuksen kokoaminen.....	25
4.1.1 Erilaisia korpustyyppisiä.....	27
4.1.2 Korpus ja käännöstutkimus	28
4.2 Korpustutkimus menetelmänä	29
4.3 Korpustutkimuksen metodit käännösten tutkimuksessa	30
5. TUTKIMUSMATERIAALI JA METODIT.....	33
5.1 Aineisto.....	33
5.2 Korpukset	34
5.3 Tutkimuskysymys ja -metodit	35
6. ANALYYSI	37
6.1 Yleistä korpuksista: sana- ja virkepituuksia	37
6.1.1 EU-korpus	37
6.1.2 SU-korpus	39
6.2 Sana-sane -suhteet sekä standardoidut sana-sane -suhteet.....	41
6.3 Sanaluokajakauma ja leksikaalinen tiheys.....	48
7. PÄÄTELMÄT.....	51

8. LOPUKSI.....	55
LÄHTEET.....	58
1.1. Kirjallisuuslähteet.....	58
1.2. Muut lähteet.....	61

LIITTEET

ENGLISH SUMMARY

1. JOHDANTO

Käännösuniversaaliteoriasta tuli suosittu tutkimuksenaihe 1990-luvulla, ja sen pohjalta toteutettiin runsaasti tutkimuksia, joissa tutkittiin käännösten kieltä korpusmenetelmiä käyttämällä. Samana vuosikymmenenä Suomi liittyi myös Euroopan unioniin, jonka kautta erityisesti institutionaalisesta kääntämisestä tuli yhä suurempi osa Suomen lainsäädäntöä. Jostain syystä nämä kaksi näkökulmaa – käännösuniversaalit ja institutionaalinen kääntäminen – eivät ole kuitenkaan kohdanneet. Tutkimukseni tarkoitus onkin tehdä juuri tämä: tarkastella institutionaalisia käännöksiä käännösuniversaalitutkimuksen näkökulmasta.

1.1 Tutkimuksen aihe ja tutkimuskysymys

Käännösuniversaaliteoriaa on tutkittu laajalti esimerkiksi kaunokirjallisuudesta koostuvalla aineistolla, mutta jostain syystä sitä ei ole ennen juuri sovellettu erikoisalojen tekstien tutkimiseen, vaikka erityisesti Euroopan unioniin liittymisen myötä lain ja hallinnon alan tekstien käännöksistä on tullut kasvava osa suomalaisten elämää ja Suomen lainsäädäntöä. Vaikka sekä institutionaalisia tekstejä ja niiden käännöksiä että alkuperäissuomen ja käännössuomen eroja onkin tutkittu laajaltikin, on aikaisempaa tutkimusta institutionaalisten tekstien suomenkielisten käännösten ja vastaavien alun perin suomeksi kirjoitettujen tekstien välisistä suhteista niukasti. Tästä syystä tutkielmani tarjoaa uuden näkökulman sekä Euroopan unionin teksteihin että käännösuniversaalihypoteeseihin yleensä.

Tutkielmassa vertailen kielellisiä eroja institutionaalisten tekstien käännöksissä käytetyn suomen ja alun perin suomeksi kirjoitettujen institutionaalisten tekstien välillä. Tutkielmani keskittyy selvittämään, mitä mahdollisia kielellisiä eroja ja samankaltaisuuksia alkuperäissuomeksi kirjoitetuilla institutionaalisilla teksteillä ja suomennetuilla institutionaalisilla teksteillä on. Keskityn tutkimaan erityisesti yksinkertaistumishypoteesia siksi, että yksinkertaistumishypoteesi tuntuu olevan ristiriidassa yleisen käsityksen kanssa, jonka mukaan Euroopan unionin suomenkieliset käännökset ovat vaikeaselkoisia ja yleisesti vaikealukuisia. Tutkin, päteeö käännösuniversaaliteorian yksinkertaistumishypoteesi myös institutionaalisissa tekstilajeissa, vai aiheuttaako institutionaalisten tekstien erikoislaatuinen luonne ja tarkoitusperä sen, että

käännöskielineen ja alkuperäissuomalainen institutionaalinen tekstien kielillä on enemmän kielellisiä yhtäläisyyksiä kuin erilaisuuksia.

1.2 Tutkielman rakenne

Luku 2 käsittelee instituution käsitettä käännöstieteellisessä kontekstissa sekä sitä, minkälaisia yleisiä piirteitä institutionaalisilla teksteillä on. Tämän lisäksi luvussa hahmotetaan Euroopan unionin ja Suomen lainsäädäntäprosesseja pääpiirteittäin ja selvitetään, miten institutionaaliset tekstit oikeastaan syntyvät, mitkä niiden tyypilliset piirteet ovat, ja mistä nämä piirteet johtuvat.

Luku 3 käsittelee deskriptiivista käännöstutkimusta sekä käännösuniversaaliteoriaa. Tämän lisäksi esittelen luvussa käännössuomen aikaisempaa tutkimusta ja sen pohjalta esitettyjä käännösuniversaalihypoteeseja. Lisäksi esittelen luvussa käännösuniversaaliteorian kritiikkiä ja pohdin, mikä tekee käännösuniversaaliteoriasta mielenkiintoisen näkökohdan tutkimukselle.

Luvussa 4 esittelen korpustutkimusta tutkimusmetodina. Kuvailen myös, miten korpustutkimuksen eri metodeilla voidaan tutkia käännöskieltä. Lisäksi esittelen erilaisia korpustutkimuksessa käytettyjä menetelmiä ja tapoja käyttää korpusta, korpustutkimuksen yleisiä vaatimuksia ja ongelmia sekä sitä, miksi juuri korpustutkimus sopii käännösuniversaaliteorian tarkasteluun.

Luvussa 5 esittelen tutkimusmateriaalini ja valintaprosessin, jota kautta olen valinnut korpuksessa olevat tekstit suhteessa luvussa 4 esitettyihin korpusta koskeviin kriteereihin. Lisäksi esittelen luvussa näistä teksteistä kootut korpuksat, niiden läpikäymän ennakoeditointiprosessin ja niiden analysointiin tutkimuksessa käytetyt työkalut.

Luku 6 koostuu aineiston pohjalta tehdystä analyysistä, jossa vertailen käännössuomen ja alkuperäissuomen korpuksen välisiä eroja tutkimalla korpuksien välisiä eroja kvantitatiivisin menetelmin yksinkertaistumishypoteesin näkökulmasta. Vertailen muun muassa korpuksien sana- ja virkepituuksia, sana-sana -suhteita sekä sanaluokkajakaumia tarkastelemalla. Luvussa 7 saatuja tuloksia analysoidaan sekä käännösuniversaaliteorian että institutionaalisen tekstien ja niiden

kääntämisen näkökulmista, ja luvussa 8 puolestaan pohdin tarkemmin saatuja tuloksia ja mahdollisia jatkotutkimusmahdollisuuksia.

2. INSTITUUTIO, SEN TEKSTIT JA NIIDEN KÄÄNTÄMINEN

Tässä luvussa keskityn hahmottamaan tarkemmin instituution käsitettä sekä tarkentamaan sitä, miten se on käännöstieteen näkökulmasta mielletty. Lisäksi tarkastelen Euroopan unionin sekä Suomen lainsäädännön tekstejä institutionaalisen kääntämisen näkökulmasta ja kuvailen näiden tekstien laadintaprosesseja pääpiirteittäin. Luonnehdin myös lyhyesti institutionaalisten tekstien erityispiirteitä.

2.1 Instituutio käännöstieteen näkökulmasta

Instituutio on laaja ja monipuolinen käsite, jolla on useita eri määritelmiä kontekstista ja tieteenalasta riippuen. Geoffrey M. Hodgsonin mukaan termiä on käytetty yhteiskuntatieteissä 1700-luvulta lähtien, mutta se on saavuttanut uutta suosiota ja sitä käytetään nykyään myös filosofian, sosiologian, politiikantutkimuksen ja maantieteen tieteenaloilla. Tästä huolimatta, sillä ei vielä ole yleisesti hyväksyttyä määritelmää. (Hodgson 2006, 1.) Sanan *instituutio* juuret ovat latinan sanassa *īstituere*, jonka merkitys on *järjestää, perustaa tai tehdä jostakin tapa tai käytäntö*, ja sanan monimerkisyys näkyy myös nykypäivänä sanan useissa eri määritelmissä. *The Cambridge Dictionary of Sociology* mukaan arkisessa kielenkäytössä sanalla *institution* viitataan heterogeeniseen joukkoon sosiaalisia normeja, kuten *perhe, kirkko* tai *monarkia*, mutta sen toisessa merkityksessä instituutioilla viitataan säännölliseen käyttäytymismalliin, joita määrittävät normit ja pakotteet, joihin yksilöt yhteiskunnassa sopeutetaan. Tästä näkökulmasta instituutiot voidaankin määritellä joukoksi sosiaalisia normeja. Perinteisessä sosiologiassa tunnustetaankin viisi erilaista yhteiskunnallisten instituutioiden joukkoa: taloudelliset instituutiot, poliittiset instituutiot, sosiaalisten kerrostumien instituutiot, perheinstituutiot sekä kulttuuriset instituutiot. (The Cambridge Dictionary of Sociology 2006, 300.) Instituutiot voidaan siis käsittää joko konkreettisina tai abstrakteina konsepteina. Esimerkiksi tietyt poliittiset, hallinnon tai sosiaalihuollon instituutiot, kuten tuomioistuin tai valtion virastot, ovat konkreettisia instituutioita, sillä niillä on konkreettista valtaa yhteiskunnassa. On kuitenkin myös abstrakteja instituutioita, kuten perheen käsite tai kulttuuriset instituutiot, joiden vaikutus ja asema yhteiskunnassa on sidottu sosiaalisiin normeihin. Näillä instituutioilla ei ole laillista valtaa, mutta niillä on tästä huolimatta huomattava asema yhteiskunnassa, mikä tekee niistä instituutioita sanan laajemmassa merkityksessä.

Myös käännöstieteen piirissä instituution ja institutionaalisen kääntämisen käsitteillä on useita eri määritelmiä. Esimerkiksi Theo Hermansin mukaan kääntäminen itsessään voidaan luokitella kulttuurilliseksi instituutioksi, jonka jokainen voi oman kulttuurinsa kontekstissa tunnistaa, ja jonka ympärille on rakentunut konkreettisia toimijoita sekä toimintatapoja (Hermans 1997, 5). Perinteisen sosiologian näkökulmasta hän siis luokittelee kääntämisen itsessään kulttuuriseksi instituutioksi. Brian Mossopin mukaan taas suurin osa kirjallisesta kääntämisestä tapahtuu jonkin instituution piirissä, sillä termin laajemmassa merkityksessä myös hallinnollisista toiminnoista irrallisina toimivat sosiaaliset rakenteet, kuten käännöksiä tuottavat kustantamot, voidaan mieltää instituutioksi. Toisin kuin esimerkiksi tulkkaminen, jossa osapuolina voivat Mossopin mukaan olla kaksi yksityishenkilöä, kuten maahanmuuttajaperheen lapsi ja tämän vanhemmat, on kääntämisen kohteena harvemmin henkilökohtainen kirjeenvaihto tai muistiinpanot. (Mossop 1988, 69–70.) Institutionaalinen kääntäminen voidaan siis käsittää hyvinkin laaja-alaisena konseptina.

Kun käsitellään konkreettisia instituutioita, kuten Euroopan unionia, on kuitenkin tarkoituksenmukaisempaa määritellä instituutionaalinen kääntäminen tämän konkreettisen instituution kontekstissa. Näin ollen tämän tutkielman aihepiirin kannalta mielekkäin näkökulma instituutionaaliseen kääntämiseen on Kaisa Koskisen määritelmä: Koskisen mukaan instituutionaalinen kääntäminen on kääntämistä, jossa virallinen taho, kuten viranomainen tai muu vastaava toimija, kommunikoi käännösten välityksellä jonkin toisen tahon kanssa. Koskisen mukaan institutionaalisessa käännöksessä kuuluu kääntävän instituution ääni, ja käännettävänä on instituutio itse. Esimerkkeinä institutionaalisen kääntämisen piiriin kuuluvista teksteistä hän mainitsee muun muassa kaksi- ja monikielisten maiden viranomaisten tuottamat viralliset dokumentit sekä Euroopan unionin ja Yhdistyneiden kansakuntien dokumenttien käännökset. (Koskinen 2008, 22.)

2.2 Instituutionaalisten tekstien erityispiirteet ja kääntäminen

Julkishallinto perustuu yhteiseen päättämiseen ja lakiteksteissä määrätään, missä ja millä tavoin asioista sovitaan (Hiidenmaa 2000, 35). Näin ollen lakitekstejä käytetään julkisten tehtävien hoitamiseen. Lakikieli on luonteeltaan normatiivista ja sen tärkein funktio yhteiskunnassa onkin ihmisten ja ihmisten välisten suhteiden määrittäminen (Cao 2007, 13). Tästä syystä se on sävyltään virallista ja persoonatonta, ja se perusominaisuuksiin kuuluu pyrkimys täsmällisyyteen ja

yksiselitteisyyteen (Koivusalo ja Huovinen-Nyberg 1985, 11). Tämän lisäksi lain ja hallinnon teksteillä on oma terminologiansa: instituutionaalisista tekstilajeista erityisesti lakikielellä on oma, oikeusjärjestelmästä riippuva sanastonsa, sekä ilmaisukäytäntöjä, jotka ovat juuri sille tyypillisiä: esimerkiksi lakienglannissa käytetään paljon vanhahtavia sanoja sekä tyypillisiä ilmauksia niiden harvinaisemmassa merkityksessä (Cao 2007, 20–21). Näin ollen lain ja hallinnon teksteissä käytetty kieli voidaan määritellä erikoiskieleksi Esko Koivusen ja Liisa Huovinen-Nybergin käyttämän määritelmän mukaan, sillä lakiteksti on tietyllä erikoisalalla käytettyä kieltä, jolla on omat, alan käyttöön vakiintuneet yleiset termit, jonka lisäksi siinä on myös usein käytössä yleiskielestä poikkeavia virkerakenteita (1985, 16–17). Toisaalta he mainitsevat samaisessa teoksessa, että juuri oikeuslaitoksen kieli on sanastoltaan yleiskielistä, mutta rakenteeltaan vaikeaselkoista (emt., 18). Myös oikeusministeriö määrittelee lakikielen kielenä, jonka pohjana on hyvä yleiskieli, mutta jonka erityispiirteitä ovat muun muassa täsmällisyys, yksitulkintaisuus sekä tiettyjen vakioilmaisuiden käyttäminen (Lainkirjoittajan opas 2013, 439).

Julkishallinnon tekstit voivat olla poikkeavia myös rakenteeltaan niin kielellisesti kuin tyylillisestikin. Esimerkiksi Pirjo Hiidenmaan tekemän tilastollisen analyysin mukaan hallinnon teksteissä käytettiin yleiskieleen verrattuna huomattavasti enemmän substantiiveja ja vähemmän verbejä, jonka lisäksi niissä oli kohtalaisen korkea leksikaalinen tiheys – tekstin sisältösanojen, kuten verbien ja substantiivien, suhde kaikkiin tekstin sanoihin oli siis jokseenkin suuri, mikä taas tekee tekstistä vaikealukuisempaa (Hiidenmaa 2000, 44–46). Hiidenmaan mukaan instituutiot eivät myöskään mukaudu sille saapuvaan ja sen ulkopuolelta tulevaan kommunikaatioon, vaan instituutio käyttää omaa esitystapaansa myös kommunikoidessaan instituution ulkopuolisten toimijoiden kanssa (Hiidenmaa 2000, 39).

Juuri nämä erityispiirteet, kuten eri terminologia sekä poikkeavat virkerakenteet ja muut yleiskielestä poikkeavat piirteet tekevät lain ja hallinnon tekstien kääntämisestä vaikeaa. Tämän lisäksi kuitenkin vaikeuksia aiheuttavat eri kulttuurien, kielten ja lakien väliset systeemiset eriävyydet. (Cao 2007, 23.) Koska lain normit sekä lakikonseptit vaihtelevat, onkin yksi lakitekstien kääntäjän haasteista lakijärjestelmien epäyhdenmukaisuus (emt., 24–25). Tätä kautta ekvivalenssin kysymys nousee esille: Koskisen mukaan ekvivalenssi (yli)korostuu instituutionaalisessa kääntämisessä, sillä instituutionaalisessa kääntämisessä eri kieliversioiden täytyy näyttäytyä ”samana” tekstinä (Koskinen 2011, 58). Tämä puolestaan on Koskisen mukaan syy siihen, että institutionaalisten tekstien käännökset tuntuvat ”oudoilta”, sillä eri kieliversioiden välinen ekvivalenssi on instituutiossa

tärkeämpää kuin kohdetekstin ja kohdekulttuurin kielisten tekstien välinen ekvivalenssi. (Koskinen 2011, 58).

Institutionaalisissa teksteissä on myös useita konkreettisia viittauksia muihin instituution tuottamiin julkishallinnon teksteihin. Nämä tekstien väliset yhteydet muihin asiakirjoihin ovat hyvin konkreettisia ja ne on usein ilmaistu selvästi viranomaisasiakirjoissa, esimerkiksi kokouspöytäkirja saattaa viitata esityslistaan ja liitteisiin. Hiidenmaan mukaan myös tämä piirre on yksi syy sille, miksi nämä tekstit ovat hallinnollisiin teksteihin tottumattoman henkilön näkökulmasta vaikeaselkoisia: ymmärtääkseen yhtä tekstiä lukijan on pystyttävä hahmottamaan sen rooli suhteessa kaikkiin muihin instituution teksteihin ja ymmärrettävä, mihin prosesseihin eri tekstit kuuluvat ja millaisessa suhteessa ne ovat toisiin teksteihin nähden. (Hiidenmaa 2000, 39.)

Merja Koskela on puolestaan tutkinut juuri tätä institutionaalisten tekstien intertekstuaalista luonnetta. Koskelan mukaan jokaisella instituutiolla on käytössä erilaisia tekstejä ja tekstilajeja. Nämä tekstit ja tekstilajit muodostavat hänen mukaansa kokonaisuuksia, jotka puolestaan heijastavat instituution toimintaa ja sen rakennetta. Koskela nimittääkin näitä tekstien ja tekstilajien välisiä suhteita tekstilajiketjuiksi. (Koskela 2007, 134.) Tämän lisäksi lain ja hallinnon tekstit ovat aina yhteydessä siihen kulttuurilliseen ja institutionaaliseen järjestelmään, johon ne kuuluvat: ne ovat siis systeemeistä riippuvaisia (Cao 2007, 23). Instituutiot siis tuottavat tekstejä, jotka peilaavat instituution toimintaa sekä sitä laajempaa yhteiskunnallista järjestelmää, jossa instituutiot toimivat.

Tämän toiminnan kuvaaminen ja selostaminen ovatkin yksi lakikielen tärkeimpiä ominaisuuksia, sillä lakikieli itsessään normatiivista ja performatiivista: tämä tarkoittaa, että se määrittää yhteiskunnallisia normeja, joiden mukaan yhteiskunnassa tulee toimia, mutta se myös toteuttaa näitä normeja kielen kautta. Sana tai lause saattaa siis jo itsessään olla teko: esimerkiksi lauseella ”Vastaaaja tuomitaan sakkorangaistukseen” on jo itsessään laillinen merkitys. (Cao 2007, 13–14.) Tai, kuten oikeusministeriön *Lainkirjoittajan opas* asian ilmaisee: ”Laki ei liioin ole ohje, ei suosituskaan. Laki sallii, oikeuttaa tai velvoittaa.” (Lainkirjoittajan opas 2013, 441). Toisin sanoen julkisissa ja institutionaalisissa teksteissä niiden eri toiminnot ja päätökset konkretisoituvat, ja nämä tekstit toimivat julkisen toiminnan areenana. Hiidenmaa kiteyttää tämän osuvasti: ”pääösteksti ei kerro vain mitä on päätetty, vaan se itsessään on päätös” (2000, 35).

Mutta miksi lain ja hallinnon tekstejä käännetään? Usein julkishallinnon tekstien kääntämisen pohjana on yksittäisten maiden lainsäädäntö itsessään, jolla taas on esimerkiksi historiallinen tai kulttuurinen syy (esimerkiksi kielilaki ja perustuslaki Suomessa, joka takaa ruotsin aseman Suomessa, sekä Official Languages Act Kanadassa). Tämän lisäksi Deborah Cao huomauttaa, että lakikääntämisen avulla on myös tuotu vieraita lakikonsepteja ja -käytäntöjä kulttuurista toiseen (2007, 32). Tämä pitää paikkansa myös esimerkiksi Suomessa: lain ja hallinnon kieli oli Suomessakin pitkään ruotsi. Mutta miksi kansainväliset instituutiot käännättävät näitä tekstejä? Syy voi olla poliittinen: Kaisa Koskisen mukaan monikielisyys nähdään positiivisena asiana, ja monikielisiä toimintatapoja pidetään osallistavina ja demokraattisina (2014, 486). Syy voi olla myös ideologinen: monikielisyyden ja -kulttuurisuuden kunnioitus on mainittu Euroopan perusoikeuskirjan 22 artiklassa (Euroopan perusoikeuskirja). Kaisa Koskinen esittää kuitenkin artikkelissaan myös konkreettisemmän syyn sille, miksi instituutiot kääntävät: ne toteuttavat kääntämällä tehtäväänsä – siis hallitsevat – kääntämisen kautta monikielisessä ympäristössään (2014, 481). Näin ollen institutionaalinen kääntäminen voidaan nähdä myös vallan käyttönä ja hallintoon liittyvänä piirteenä monikielisissä yhteisöissä.

2.3 Suomen lainsäädäntö ja lainvalmisteluprosessi

Suomessa lainsäädäntövaltaa käyttää Eduskunta, joka käsittelee lakiehdotuksia täysistunnossa. Vaikka lakiehdotuksen voi tehdä yksittäinen kansanedustaja ja vuodesta 2012 lähtien sen on voinut tehdä myös kansalaisaloitteena, käytännössä suurin osa voimaantulevista lakialoitteista ja lakimuutoksista on lähtöisin hallituksen esityksistä (Eduskunta). Lainvalmisteluprosessi voi käynnistyä myös muun muassa Eduskunnan lausuman, lakimuutosta vaativan valtion talousarvioehdotuksen, EU-tuomioistuimen lausuman tai komission tulkinnan tai EU-säädöksen laatimisen alkamisesta. Tämän jälkeen lakiesitys käy läpi seitsemän eri vaihetta: esivalmisteluvaiheen, perusvalmisteluvaiheen, lausuntomenettelyn, jatkovalmistelun, valtioneuvoston päätöksentekovaiheen, eduskuntakäsittelyyn sekä lain vahvistamisen, jonka jälkeen vahvistettu laki julkaistaan Suomen säädöskokoelmassa ja se tulee voimaan laissa mainittuna päivänä. (Oikeusministeriö).

Tämän säädäntäprosessin aikana lukuiset eri osapuolet laativat erilaisia asiakirjoja, kuten muistioita, raportteja ja kannanottoja, jotka ovat kaikki osa lainsäädäntäprosessia ja vaikuttavat prosessin lopputuloksena syntyvään lakiin. Nämä eri tekstilajit ovat osa prosessia, jossa poliittinen instituutio, siis Eduskunta, toteuttaa mandaattiaan eli lainsäädäntävaltaansa. Nämä eri tekstilajit puolestaan heijastavat kyseisen instituution toimintaa ja toimintatapoja. Tästä syystä voidaankin puhua Merja Koskelan termiä käyttäen *tekstilajiketjuista*, jotka perustuvat Koskelan mukaan ”instituution vakiintuneisiin tapoihin hoitaa tehtäviään eri tekstien kautta” (Koskela 2007, 134).

On myös huomioitava, että useat eri osapuolet sekä yksilöt osallistuvat näiden dokumenttien laatimiseen ja kirjoittamiseen, ja lopullinen laki pohjautuukin siis sitä edeltäneisiin teksteihin. Näin ollen lakitekstin kirjoittaminen voidaan nähdä olevan eräänlaisen instituution sisäisen yhteiskirjoittamisen tulos.

2.3.1 Lainsäädännön prosessit

Esivalmisteluvaiheessa kootaan tietoa ja näkemyksiä lakia koskevasta aiheesta joko virkatyönä tai sitä varten asetetussa valmisteluelimessä. Aluksi määritellään tavoite tai ongelma, jota varten uusi laki tai lakimuutos tulisi säätää, ja tämän jälkeen tehdään lakiesityksen aiheesta alustava kartoitus. Tämän jälkeen valmistelija tai valmistelutiimi esittää näkemyksensä erilaisille lakiesitystä koskeville viranomaisille, asiantuntijoille ja sidosryhmille, selvittää niiden näkemyksen lakiesityksestä ja hankkii niiltä lisätietoja lakiesityksen aiheesta, jonka jälkeen virkamiesjohto tekee päätöksen esivalmistelun käynnistämisestä. Esivalmistelussa selvitetään laajemmin esimerkiksi aihetta koskevia tutkimuksia, oikeuskäytäntöjä sekä kirjallisuutta. Esivalmisteluvaiheen jälkeen lakiehdotus siirtyy perusvalmisteluun. (Oikeusministeriö.)

Perusvalmisteluvaihe on yleisimmin lainvalmistelun eniten työpanosta vaativa vaihe, jossa tarkennetaan ja laajennetaan esivalmisteluvaiheessa koottuja tietoja ja selvitetään lakiesityksen suhdetta perustuslakiin ja jo olemassa olevaan lainsäädäntöön sekä konsultoidaan asiantuntijoita ja sidosryhmiä. Tämä voidaan tehdä joko virkatyönä, jolloin prosessista vastaa valmistelutiimi, tai valmisteluelimessä, mikäli valmistelu tehdään laajempaan kokonaisuuteen kuin virkamiestyönä. Mikäli valmistelu tehdään virkamiestyönä, prosessin lopuksi valmistelutiimi tekee ehdotukset

sisällön linjauksesta, jonka ministeri, hallitus tai virkamiesjohto käsittelee lain merkittävyydestä riippuen. Mikäli linjaukset hyväksytään ja prosessissa päädytään etenemään hallituksen esityksen luonnoksen (HE-luonnoksen) laadintaan, kirjoittaa valmistelutiimi sisällön linjausten perusteella lakiluonnoksen sekä sen perustelut HE-luonnoksen muotoon. HE-luonnoksen hyväksymisestä päättää joko ministeri, hallitus tai virkamiesjohto, jonka jälkeen laki siirtyy lausuntomenettelyyn. Mikäli prosessista vastaa valmistelualue, päättää se itse puheenjohtajan johdolla luonnoksen sisällöstä sekä HE-luonnoksen hyväksymisestä. Kun valmistelualue on hyväksynyt HE-luonnoksen, laaditaan loppuraportti, joka luovutetaan ministeriölle, ja säädösehdotukset sekä tiivistelmä käännetään ruotsiksi. Tämän jälkeen laki siirtyy lausuntomenettelyyn. (Oikeusministeriö.)

Kun luonnos hallituksen esitykseksi on hyväksytty perusvalmisteluvaiheessa, pyydetään siitä kirjallista lausuntoa sidosryhmiltä, joille lähetetään lausuntopyyntö ja tarvittaessa kutsu kuulemistilaisuuteen. Lisäksi HE-luonnos käännetään ruotsiksi, ellei sitä ole perusvalmisteluvaiheessa vielä käännetty. Kun sidosryhmät ovat antaneet lausuntonsa ja niissä annettu palaute on käsitelty, siirtyy laki jatkovalmisteluun, jossa sidosryhmille annetaan palaute lausunnoista ja ne osallistuvat keskustelu- tai neuvottelutilaisuuteen, jonka pohjalta tehdään päätös HE-luonnoksen sisällön muuttamisesta tai sen hyväksymisen käsittelystä. Tämän jälkeen HE-luonnos viimeistellään, tarkastetaan ja käännetään uudelleen, ja oikeusministeriön laintarkastusyksikkö tarkastaa ne. Lopuksi valmistelijatiimi laatii ja jakaa esityslistat valtioneuvostolle. (Oikeusministeriö.)

Valtioneuvoston yleisistunnossa hallitus päättää, annetaanko esitys eduskunnan käsiteltäväksi. Mikäli esitys siirtyy eduskunnan käsiteltäväksi, toimittaa valtioneuvoston istuntoyksikkö sen allekirjoitettuna eduskunnalle ja valmistelutiimi antaa esityksestä painatusmääräyksen. Painotalo toimittaa painetun esityksen eduskunnalle. Eduskunta päättää esityksen hyväksymisestä tai hylkäämisestä; hyväksytty laki esitetään presidentin vahvistettavaksi. Jos presidentti ei vahvista lakia allekirjoituksella palautuu se eduskunnan käsiteltäväksi. Lopuksi valtioneuvoston istuntoyksikkö laatii pöytäkirjat istunnosta ja presidentin päätöksestä. Hyväksytty laki lähetetään julkaistavaksi Suomen säädöskokoelmaan, ja se astuu voimaan laissa mainittuna päivänä. (Oikeusministeriö.)

2.3.2 Kielelliset standardit ja kääntäminen osana Suomen lainsäädäntöprosessia

Kuten Suomen kielilaki 6.6.2003/423 selvästi ilmaisee, ovat Suomen kansalliskielet suomi ja ruotsi (1 luku § 1), ja lait säädetään ja julkaistaan sekä suomen että ruotsin kielellä (7 luku § 30). Tämän lisäksi Suomen kansalliskielistä on maininta Suomen perustuslaissa 11.6.1999/731, jossa määritellään, että jokaisella on oikeus omaan kieleen ja kulttuuriin (luku 2 § 17). Tämän lisäksi Suomen Hallintolaissa 6.6.2003/434 säädetään, että viranomaista sitoo niin sanottu hyvän kielen vaatimus, jonka mukaan viranomaisella on velvollisuus käyttää asiallista, selkeää ja ymmärrettävää kieltä (luku 2 § 9).

Tämän lainsäädännön perusteella voisi siis päätellä, että Suomessa lainsäädäntö tehdään samanaikaisesti suomen ja ruotsin kielellä. Näin ei kuitenkaan käytännössä ole. Kielilain mukaan valtion virkamiehien työkieli on virka-alueen väestön enemmistön kieli, ellei perustellusta syystä muun kielen käyttö ole tarkoituksenmukaisempaa (6 luku § 26). Näin ollen suomen lainsäädäntö siis käytännössä kirjoitetaan suomeksi ja käännetään ruotsiksi.

Käytännössä eri lainvalmisteluprosessiin liittyviä tekstejä käännetään kuitenkin lähes koko prosessin ajan. Lain perusvalmisteluvaiheen loppuvaiheessa valmistuneet säädösehdotukset sekä perusvalmisteluvaiheessa lakiesityksen vaikutuksista laadittu tiivistelmä käännetään ruotsiksi, ja lausuntomenettelyvaiheessa ruotsiksi käännetään hallituksen esityksen luonnos (HE-luonnos) sekä lausuntopyyntö. Jatkovalmistelussa taas viimeistelty hallituksen esityksen luonnos käännetään ruotsiksi ja sen vastaavuus suomenkielisen tekstin kanssa tarkastetaan. (Oikeusministeriö.) On kuitenkin selvää, että vaikka tekstejä käännetään useamman kerran prosessin eri vaiheissa, on käännös yleensä jokaisen vaiheen lopussa tuotettu teksti: teksti on siis jo jonkin prosessin tuloksena syntynyt ”valmis” teksti siinä vaiheessa, kun se käännetään ruotsiksi. Poikkeuksena kuitenkin on virkamiestyönä tehty perusvalmisteluvaihe, jossa kääntäjille saatetaan hallituksen esityksen luonnoksen valmisteluvaiheessa lähettää lista termeistä ja nimistä selityksineen, ja niiden käännettävyydestä pyydetään palautetta – tosin näin tapahtuu vain ”tarvittaessa” (Oikeusministeriö).

2.4 Euroopan unionin lainsäädäntö

Euroopan unionissa päätöksenteko tapahtuu useiden erilaisten oikeudellisten säädösten kautta, joiden sitovuus vaihtelee säädöksestä toiseen. Tämä päätösvalta perustuu jäsenmaiden sopimiin perussopimukseen, joiden ulkopuolelle jäävistä politiikanaloista ei voida Euroopan unionissa laatia lakiehdotuksia (Ruonala & Pietilä 2014, 60).

Euroopan unionin sitovia päätöksiä ovat asetukset, direktiivit sekä päätökset. Asetukset ovat sitovia määräyksiä, jotka koskevat kaikkia jäsenmaita ja jonka kaikkia osia on sovellettava kaikkialla Euroopan unionin alueella. Direktiivit ovat myös sitovia säädöksiä, mutta ne määrittelevät vain sitovan tavoitteen, johon kaikkien EU-maiden tulee pyrkiä. (Ruonala & Pietilä 2014, 63.) Käytännössä tämä tarkoittaa sitä, että jäsenmailla on kansallisessa lainsäädännössään vapaus määrätä siitä, miten direktiivin määrittämä sitova määräys käytännössä saavutetaan. Päätökset ovat puolestaan sitovia päätöksiä, jotka koskevat niitä osapuolia, kuten jäsenmaita tai yrityksiä, joille ne on osoitettu. Näiden sitovien päätösten lisäksi Euroopan unionin elimet voivat antaa suosituksia tai lausuntoja, jotka eivät ole sitovia. (Europa).

2.4.1 Euroopan unionin lainsäädäntöprosessi

Euroopan unionissa tavallisessa lainsäätämisenjärjestelymenettelyssä komissiolla on aloiteoikeus tehdä lakiehdotuksia, joita se tekee useimmiten Euroopan unionin perussopimuksen velvoittamana. Tämän lisäksi muut osapuolet, kuten sidosryhmät, jäsenmaat, EU:n muut toimielimet tai vuodesta 2012 lähtien Euroopan unionin kansalaiset eurooppalaisen kansalaisaloitteen muodossa voivat kuitenkin pyytää komissiota laatimaan lakialoitteen (Europarl). Lakialoitteesta konsultoidaan useita osapuolia, kuten sidosryhmiä, jäsenmaita, ja tarvittaessa myös EU:n muita toimielimiä, jonka jälkeen lakiesitys siirtyy Euroopan unionin parlamentin ja niissä toimivien komiteoiden käsiteltäväksi (Europarl), jolloin se käännetään kaikille unionin virallisille kielille (Robinson 2012, 10). Lakialoite on usein samanaikaisesti myös Euroopan unionin neuvoston käsittelyssä, joskin se voi antaa oman kantansa virallisesti vasta parlamentin kannan saatuaan. Mikäli neuvosto haluaa muuttaa parlamentin kantaa, palautuu lakialoitteen käsittely parlamentin toista käsittelyä varten, josta parlamentin mahdollisesti tekemät tarkastukset siirtyvät taas neuvoston hyväksyttäväksi. (Europarl.)

Jos neuvosto hyväksyy kaikki parlamentin tarkistukset, tulee säädös voimaan. Mikäli neuvosto ei hyväksy parlamentin tarkastuksia sellaisenaan, kutsuvat neuvoston puheenjohtaja ja Euroopan parlamentin puhemies kokoon sovittelukomitean, johon kuuluu Euroopan parlamentin sekä neuvoston edustajia. Mikäli sovittelukomitea pääsee yhteisymmärrykseen yhteisestä tekstistä, siirtyy teksti parlamentin täysistunnon ja neuvoston kolmanteen käsittelyyn. Jos yhteisymmärrykseen ei päästä, säädösesitys raukeaa. Säädösmenettely päättyy myös, mikäli joko parlamentin täysistunto tai neuvosto ei hyväksy uutta tekstiä. Uusi teksti on hyväksyttävä sekä parlamentin täysistunnossa että neuvostossa sellaisenaan, eikä siihen voida enää tehdä muutoksia. (Europarl.)

2.4.2 Monikielinen lainsäädäntö ja kääntäminen osana lainsäädäntöprosessia Euroopan unionissa

Euroopan unionin pyrkii tukemaan monikielisyyttä, mikä mainitaan Euroopan unionista ja Euroopan unionin toiminnasta tehdyn yhdistetyn sopimuksen artiklassa 3:

Unioni kunnioittaa kulttuuriensa ja kielensä rikkautta ja monimuotoisuutta sekä huolehtii Euroopan kulttuuriperinnön vaalimisesta ja kehittämisestä.

Käytännössä tämä tarkoittaa sitä, että kaikki Euroopan unionin viralliset tekstit ovat saatavilla kaikilla unionin virallisilla kielillä, ja että jokaisella Euroopan unionin kansalaisella on oikeus ottaa yhteyttä Euroopan unioniin ja saada vastaus omalla äidinkielellään. Kaikilla Euroopan unionin europarlamentaarikoilla on myös oikeus tulkkaukseen. Tämän lisäksi Euroopan unioni pyrkii edistämään monikielisyyttä muun muassa tukemalla kielten oppimista ja yhteiskunnan kielellistä monimuotoisuutta (Monikielisyyden puitestrategia 2005). Tästä huolimatta komissiossa, jolla on aloiteoikeus ja jossa lakiesitykset kirjoitetaan, on kolme työkieltä: englanti, ranska sekä saksa, ja käytännössä suurin osa lakiesityksistä kirjoitetaan englanniksi (Reiman 2014). Lakiesityksiä kirjoittavat niin sanotut ”policy officerit” ovat myös harvoin lain asiantuntijoita, ja usein heidän täytyy työskennellä muulla kuin äidinkielellään (Robinson 2014, 11). Euroopan komission käännöstoimen pääosaston vuonna 2009 tekemä tutkimus tuli samaan lopputulokseen: vaikka 95 prosenttia komission lakiesityksiä kirjoittavista teki töitä englanniksi, vain 13 prosentilla englanti oli heidän pääkielensä, ja useat heistä eivät juuri luetuttaneet kirjoittamiaan tekstejä henkilöllä, joka puhuu kieltä äidinkielenään (Wagner 2010, 4).

Kaikki lakiesityksen on kuitenkin käännettävä kaikille Euroopan unionin virallisille kielille ennen kuin ne voidaan toimittaa jatkokäsittelyyn neuvostoon ja parlamentille. Periaatteessa kääntäminen tulee siis mukaan lainsäädäntömenettelyyn verrattain varhaisessa vaiheessa. Tästä huolimatta käytännön syistä suurin osa keskusteluista rajataan koskemaan vain yhtä kieliversiota lakiesityksestä, joka on yleensä englanninkielinen versio, mutta joskus myös ranskankielinen versio. (Robinson 2009, 10–11.) Lainsäädäntövaiheen loppuvaiheessa, ennen parlamentin täysistuntoa, sekä komission että parlamentin laki-lingvistit käyvät läpi lakiesityksen yhdessä komission, neuvoston ja parlamentin asiantuntijoiden kanssa. Tämän prosessin tuloksena syntyy yksi pohjateksti, joka on yleensä englanninkielinen. Tämä pohjateksti taas käydään läpi niin sanotussa laki-lingvistien kokouksessa. Kokouksessa koko teksti käydään läpi, ja tämän jälkeen teksti jaetaan muiden kielten laki-lingvisteille. Tässä vaiheessa onkin vaikeaa enää parantaa tekstin laatua, sillä se saattaisi johtaa tekstiin kirjoitettujen poliittisten kompromissien rikkoutumiseen. Tämän lisäksi kokouksessa läsnä olevat jäsenmaiden edustajat voivat vastustaa tehtyjä muutoksia. Kokouksen jälkeen jokaisen kielen laki-lingvistit käyvät tekstin läpi omalla kielillään ja jäsenmaiden ammattilaisten avustuksella, jonka pohjalta valmis versio lähetetään parlamentin täysistuntoon äänestystä varten. (Robinson 2009, 14.) Käytännössä kääntäminen siis ei juuri ole mukana itse lainsäätämisprosessissa, vaan kääntäminen tulee mukaan vasta lainsäädännön viimeistelyvaiheessa.

Euroopan unioni pyrkii kuitenkin ottamaan huomioon kääntämisen ja monikielisen lainsäädännön asettamat vaatimuksen lakiesityksiä kirjoittaville viranomaisille tarkoitettussa ohjeistuksessaan (Joint Practical Guide 2013). Oppaan mukaan koko lainsäädäntökaaren aikana lakiesityksen kielen tulee käyttää termejä ja kielellisiä rakenteita, jotka heijastavat Euroopan unionin monikielisyyttä. Oppaassa varoitetaan lisäksi tietyn jäsenmaan laki- tai kielikulttuurille uniikkien termien ja ilmaisuiden käytöstä sekä hankalista lauserakenteista. Tämän lisäksi oppaassa huomautetaan, että kääntäjien ja muiden osastojen kielitarkastuksesta saatu palaute on hyödyllistä, ja että useassa tapauksessa on itse asiassa parempi muuttaa alkuperäistä tekstiä eikä siitä tehtyä käännöstä. (Joint Practical Guide 2013, 11–13). Tämän lisäksi Euroopan komission käännöstoimen pääosasto on julkaissut kielioppaita, kuten *How To Write Clearly* (2012) sekä *English Style Guide* (2011, viim. päivitetty 2015).

Tämän pohjalta voidaan siis todeta, että vaikka kääntäminen prosessina ei välttämättä olekaan mukana lakiesityksen eri prosesseissa, on se otettu huomioon viranomaisille suunnatun ohjeistuksen

muodossa. Voisi myös sanoa, että ohjeistuksensa ja oppaidensa kautta Euroopan unioni yrittää aktiivisesti luoda kontrolloitua kieltä, josta se pyrkii tekemään mahdollisimman ”käännettävää”.

2.5 Yhteenveto: instituutiot ja niiden tekstit

Kääntämisellä ja käännöksillä on selkeästi merkittävä osa sekä Suomen että Euroopan unionin lainsäädännössä, mutta niitä kohdellaan hieman eri tavalla: Suomessa ruotsinkielinen käännös otetaan selvemmin osaksi prosessia ja sitä tehdään koko prosessin ajan. Euroopan unionissa taas kääntäminen tapahtuu pitkälti kahdessa vaiheessa: kun esitys ensin saapuu parlamentin ja neuvoston käsittelyyn, ja kun sen viimeisteltyä versiota valmistellaan parlamentin täysistuntoon. Kummassakin tapauksessa kääntäjät työskentelevät kuitenkin itse lainsäädäntö- ja kirjoitusprosessista erillään. Vaikka Suomessa käännetäänkin enemmän tekstejä itse prosessin aikana, ovat kyseiset tekstit ”valmiita” esityksiä ja ”valmiita” luonnoksia, ja kääntäminen tapahtuu lähes aina näiden vaiheiden loppupäässä. Sama pätee Euroopan unionissa.

Toisaalta Euroopan unionissa suurin osa säädösesityksiä kirjottavista viranomaisista ei puhu käyttämäänsä työkieltä äidinkielenään, mikä ei pidä paikkansa Suomessa. Tästä syystä Euroopan unioni on nostanut kääntämisen esille itse tekstin kirjoitusvaiheessa hyvin konkreettisella tavalla. Siis, vaikka tekstiä ei käännetäkään säädösten valmisteluvaiheessa, kannustetaan tekstin käännettävyyden mielessä pitämistä prosessin alkuvaiheesta lähtien. Sama pätee myös hieman vähemmässä määrin Suomessa tapauksissa, jolloin kääntäjiltä pyydetään palautetta termien käännettävyydessä lain perusvalmisteluvaiheessa.

Kaiken kaikkiaan lainvalmistelua molemmissa lainsäädäntöjärjestelmissä yhdistää joukkovalmistelu: lakeja kirjoittaa ja niiden säätämisessä on mukana suuri joukko eri viranomaisia, sidosryhmiä ja muita työryhmiä. Ulla Tiilikä kirjoittaa, että virkatekstit laaditaan usein tiimityönä ja usein toisten tekstien pohjalta (2001, 9), ja tämä piirre yhdistää sekä Euroopan unionin että Suomen lakitekstejä. Lainsäädäntö on siis eräänlaista joukkokirjoittamista sekä Euroopan unionissa että Suomessa. Tämän lisäksi molemmilla instituutioilla on käytössä esimerkiksi lakia kirjoittaville viranomaisille laadittuja oppaita, joiden kautta kontrolloidaan niin käytettyä kieltä kuin lain muodollisia rakenteitakin (Joint Practical Guide 2013; Lainkirjoittajan opas 2013). Lisäksi molemmissa instituutioissa Koskelan

mainitsevat tekstilajiketjut näkyvät selvästi: tekstejä rakennetaan jatkuvasti jo olemassa olevien lakien pohjalle, ja itse lainsäädäntöprosessinkin aikana molemmissa instituutioissa esitykset, mietinnät ja luonnokset muodostavat merkityksellisen ketjun, jonka kautta prosessi rakentuu. Tämä näkyy esimerkiksi Euroopan unionin teksteissä hyvin konkreettisella tavalla: tekstin aluksi listataan, mitä lausuntoja ja ehdotuksia lakia laadittaessa on otettu huomioon. Esimerkiksi Euroopan parlamentin ja neuvoston asetus N:o 524/2013 (ks. liite B) alkaa seuraavasti:

EUROOPAN PARLAMENTTI JA EUROOPAN UNIONIN NEUVOSTO, jotka

ottavat huomioon Euroopan unionin toiminnasta tehdyn **sopimuksen** ja erityisesti sen 114 artiklan,

ottavat huomioon Euroopan komission **ehdotuksen**,

sen jälkeen kun **esitys** lainsäätämisyjärjestyksessä hyväksyttäväksi säädökseksi on toimitettu kansallisille parlamenteille,

ottavat huomioon Euroopan talous- ja sosiaalikomitean **lausunnon** ⁽¹⁾,

noudattavat tavallista lainsäätämisyjärjestystä ⁽²⁾,

sekä katsovat seuraavaa:

(kirjoittajan lihavointi)

Tekstit siis viittaavat suoraan niitä edeltäneisiin teksteihin, johon ne pohjautuvat, sekä komiteoihin, joita ne ovat konsultoineet ja joiden lausuntoon neuvosto ja parlamentti ovat tutustuneet ennen päätöksentekoa. Koskelan termiä käyttäen voidaankin siis sanoa, että tekstin tekstilajiketjua on mahdollista tarkastella itse tekstistä. Tekstit viittaavat selvästi instituution muihin teksteihin, ja tämä tekstien välinen intertekstuaalisuus ja muihin instituution tuottamiin teksteihin viittaaminen ovatkin viranomaiskielelle tyypillinen piirre (Hiidenmaa 2000, 39). Mutta miten monikielinen toimintaympäristö vaikuttaa näihin suhteisiin? Varsinkin kun on kyse jäsenmaita velvoittavista teksteistä, joista itsestään tulee osa Suomen lainsäädännön tekstilajiketjuja, on syytä kysyä, miten tämä suhde vaikuttaa tekstien kääntämiseen – vai vaikuttaako se lainkaan? Toisaalta, kuten yllä todettiin, on tekstien kirjoittamisprosessi hyvin samantapainen kummassakin instituutiossa. Tämän lisäksi teksteillä on keskenään samanlaisia piirteitä, sama tekstityyppi sekä hyvin samanlainen, hyvin tarkasti määritelty funktio. Jos tekstit on siis tuotettu hyvin samanlaisen prosessin tuloksena hyvin samanlaiseen tarkoitukseen, voidaanko niissä olettaa olevan samanlaisia kielellisiä piirteitä? Entä

miten nämä piirteet vaikuttavat tekstien kääntämiseen ja sen lopputuloksena syntyvään käännökseen? Tiililän mukaan tekstin työn ja tekemisen tavat vaikuttavat tekstin ominaispiirteisiin (2001, 11), joten voidaanko olettaa, että Euroopan unionin ja Suomen lainsäädännön samantapaiset tekstin tuottamisen tavat vaikuttavat siihen, millaisia Suomen lainsäädännön tekstiketjun osaksi käännetty Euroopan unionin säädökset käännetään? Tämän kysymyksen pohjalta onkin mielenkiintoista tutkia, löytyykö Euroopan unionin säädösteksteistä käännöksille tyypilliseksi luokiteltuja piirteitä.

3. KÄÄNNÖSUNIVERSAALITEORIA JA KÄÄNNÖSSUOMEN TUTKIMUS

Tässä kappaleessa käsittelen käänносuniversaaliteoriaa, sen historiaa ja siihen kohdistunutta kritiikkiä. Lisäksi käsittelen aiheen aikaisempaa tutkimuksia sekä siitä saatuja tuloksia.

3.1 Käännosuniversaaliteoria

Käännosuniversaaliteoria, jonka mukaan käännöksistä löytyy niille ominaisia kielellisiä piiteitä, on syntynyt deskriptiivisen käännostutkimuksen vaikutuksesta. Deskriptiivinen käännostutkimus (*descriptive translation studies*, DTS), jossa keskitytään tutkimaan käännostä itsessään, syntyi pitkälti vastareaktionä arvottavalle ja kääntäjien koulutukseen keskittyvälle käännostutkimukselle 1970- ja 1980-luvuilla (Hermans 2012, 1681). Tähän vaikutti se, etteivät käännökset juurikaan saaneet akateemista huomiota ennen 1900-luvun puoliväliä (Hermans 2012, 1681), eikä käännöksiä pidetty kuin ”toisen käden ’aitojen’ tekstien vääristyneinä kuvina” (Baker 1993, 233; tekijän suomennos). Myös kirjallisuudenhistoriassa mainintoja käännöksiin tehdään yleensä vain, jos se on täysin välttämätöntä (Even-Zohar 1990, 45). Deskriptiivisen käännosteorian juuret ovatkin tämän historiallisen näkökulman kritiikissä: sen pohjana on käännosten tutkiminen osana kulttuurihistoriaa sekä niiden tutkiminen kulttuurisena toimintana (Herman 2012, 1681).

Käännosuniversaaliteorian pohjana voidaan pitää Itamar Even-Zoharin kehittämää polysysteemiteoriaa sekä deskriptiivistä käännostutkimusta, jossa keskitytään tutkimaan käännostä itsessään. Termi deskriptiivinen käännostutkimus (*descriptive translation studies*) on peräisin James S. Holmesin artikkelista *The Name and Nature of Translation Studies* (2005 (1972)), jossa hän pyrki määrittelemään käännostieteen eri empiiriset tutkimusalat. Holmesin jaottelussa deskriptiivinen käännostutkimus keskittyy kuvailemaan kääntämistä sekä käännöksiä empiirisellä tasolla (2005 (1972), 71–72). Holmesin ideat innoittivat puolestaan Gideon Tourya, jota voidaan nykyään pitää käännostieteen normeja ja lainalaisuuksia koskevan tutkimuksen pioneerinä.

Itamar Even-Zoharin polysysteemiteorian mukaan semioottiset ilmiöt, kuten kieli ja kirjallisuus, tulisi nähdä heterogeenisinä systeemeinä, eikä toisistaan erillään olevina elementteinä. (1979, 287). Even-

Zohar kysyykin, onko mahdollista, että käännökset muodostavat samanlaisia keskinäisiä kulttuurillisia ja kielellisiä verkostoja, kuten eri kulttuurien alkuperäiskielisten kirjallisuuksien on esitetty tekevän (1990, 45–46). Hänen mukaansa käännöstekstit toimivat systeeminä kahdella tavalla: kun systeemit valitsevat, mitä tekstejä sekä mistä kielistä ja kulttuureista käännetään, sekä kun ne valitsevat, mitä normeja ja malleja ne käännöksissä noudattavat. Tätä kautta käännöksillä saattaakin olla oma järjestelmänsä, joka on käännöksille uniikki. (Emt. 46.)

Gideon Toury puolestaan käsittelee myös käännösten tutkimista kulttuuris-systeemisestä näkökulmasta. Kirjassaan *Descriptive Translation Studies – And Beyond* (1995), Toury argumentoi, että kääntäminen pitäisi nähdä toimintana, jolla on kulttuurista merkitystä, ja meidän tulisi tunnustaa, että kääntäjällä on näin ollen sosiaalinen rooli omassa kulttuurissaan. Tämän roolin suorittamista puolestaan määrittävät normit, joiden noudattaminen on edellytys kääntäjänä toimimiselle kyseisessä kulttuurissa. (Emt., 53.) Koska kääntäminen on toimintaa, jossa on mukana vähintään kaksi eri kieltä ja kulttuuria, ovat nämä eri kielten ja kulttuurien kääntäjälle asettamat normit Touryn mukaan, elleivät täysin yhteen sopimattomia, niin ainakin huomattavan erilaisia. Touryn mukaan käännöstoiminnassa esiintyykin kulttuurin sisäisiä säännönmukaisuuksia, jonka kautta näiden kahden normijoukon välisiä jännitteitä voidaan selvittää. (Emt., 56.) Kääntäjät siis tasapainottelevat kohdekielen ja lähdekielen normien välillä ja käännöksiä tutkimalla voidaankin selvittää, millaisia normeja he päätyvät noudattamaan, sekä millaisia normeja heidän kulttuurissaan käännöksien oletetaan noudattavan – esimerkiksi minkäläisten käännösstrategioiden käyttö koetaan suotavana tai hyväksyttävänä, ja minkäläisten strategioiden käyttö nähdään vääränä tai epätoivottavana.

Toury jakaa kääntämistä määrittävät normit ennakkonormeihin (*preliminary norms*), jotka määrittävät kääntämisen laajempia käytäntöjä, sekä toiminnallisiin normeihin (*operational norms*), jotka määrittävät käännösprosessia itseään. Ennakkonormeiksi hän määrittää muun muassa käännöspoliittiset normit – mitä tekstejä tai tekstilajeja käännetään, mistä kielistä ja mistä kulttuurista – sekä käännöksen suoruutta määrittävät normit eli se, kuinka kulttuuri suhtautuu käännöksiin, jotka on käännetty jostain muusta kielestä kuin alkuperäisen teoksen lähtökielestä. (Toury 1995, 58). Toiminnallisiksi normeiksi hän taas lukee matriaaliset normit (*matricial norms*) sekä teksti-kielelliset normit (*textual-linguistic norms*). Matriaaliset normit määrittävät lähdekielisen materiaalin korvaamaan tarkoitetun kohdekielisen materiaalin esiintymistä käännöksessä, kuten esimerkiksi sitä, millaisia poistoja ja lisäyksiä käännökseen on tehty. (Emt. 58–59.) Teksti-kielelliset normit puolestaan määrittävät sitä, millaista kielellistä materiaalia käyttämällä käännös tehdään (emt. 59),

eli yksinkertaisesti: millaisia kielellisiä ratkaisuja kääntäessä tehdään. Touryn mukaan yksi tapa tutkia käännösnormeja onkin tutkia juuri käännöksiä, sillä käännökset ovat hänen mukaansa normeihin perustuvan toiminnan suoraa tulosta (emt., 65). Tämän oletuksen pohjalta Toury päätyy ehdottamaan mahdollisia kääntämiselle tyypillisiä piirteitä, kuten kasvavan standardisoinnin lain (emt., 267), sekä interferenssin lain (emt., 274–275). Sanan laki (*law*) käyttöä hän perustelee sillä, että jokainen tieteenala etsii alansa koskevia lakeja ja säännönmukaisuuksia, eikä hänestä ole mitään syytä, miksi käännöstieteen tulisi olla poikkeus (emt., 259).

Touryn näkökulmaa kääntämisestä normien määrittämänä toimintana sekä kääntämistä koskevista laeista ovat puolestaan kehittäneet eteenpäin muun muassa Andrew Chesterman, Theo Hermans sekä Mona Baker. Vuonna 1993 julkaistussa artikkelissaan *Corpus linguistics and translation studies* Mona Baker esittää Touryn ja Even-Zoharin teorioihin pohjaten, että näiden normeja olisi mahdollista tutkia korpusta apuna käyttäen, ja Baker mainitseekin käännösten universaalit piirteet yhtenä mahdollisena tutkimuskohteena (Baker 1993, 243). Näistä piirteistä hän mainitsee muun muassa eksplisioitumisen, yksinkertaistumisen, konventionaalisen kieliopin korostumisen, toiston välttämisen, kohdekielen piirteiden ylikorostamisen, sekä kielten tiettyjen piirteiden poikkeava esiintyminen alkuperältään kohdekieliseen tekstiin verrattuna (Baker 1993, 244–245). Myöhemmässä artikkelissaan Baker kuvaili, kuinka esimerkiksi juuri korpustutkimuksesta olisi apua näitä universaaleja tutkittaessa (Baker 1995, 235; ks. luku 4.2.1).

Andrew Chesterman puolestaan lajittelee olemassa olevan tutkimuksen mukaan ehdotetut käännösuniversaalit kahteen eri luokkaan: lähdeuniversaaleihin (*source universals*, *S-universals*) sekä kohdeuniversaaleihin (*target universals*, *T-universals*) (2004, 37). Lähdeuniversaaleilla hän tarkoittaa tapaa, jolla käännökset eroavat lähdeteksteistä, ja kohdeuniversaaleilla taas käännöstekstien piirteitä, jotka erottavat ne vertailukelpoisista kohdekielisistä teksteistä (Chesterman 2004, 39). Mahdollisiksi lähdeuniversaaleiksi hän ehdottaa alan tutkimukseen nojaten muun muassa käännöksen pituuden lisääntymistä, interferenssiä, standardisaatiota, murteen normalisointia sekä eksplikointia. Mahdollisiksi kohdeuniversaaleiksi hän taas luettelee muun muassa yksinkertaistumisen, epätyypilliset leksikaaliset esiintymät sekä kohdekielen uniikkien aineiden vähäisen esiintymisen käännöksissä. (emt., 40).

3.2 Käännösuniversaaliteorian kritiikki

Käännösuniversaaliteorian suosiosta huolimatta sitä on myös kritisoitu. Ensinnäkin teoriaa koskevat hypoteesit ja niiden tutkimus on rajautunut vain nykykielen ja suppeaan kielivalikoimaan piiriin, mistä syystä on mahdotonta määrittää yleismaailmallisia – eli universaaleja – hypoteeseja (Jantunen 2004, 50). On myös ongelmallista määritellä käännösuniversaaleja, mikäli tutkittavat käännökset eivät ole edustava joukko kaikista käännöksistä (Chesterman 2004, 42–43). Myös Maria Tymoczko huomauttaa, että käännöksissä esiintyvien lainalaisuuksien etsimisestä tekee ongelmallista jo se, että käännös itsessään on kahden kulttuurin määrittämä tapahtuma, jossa sekoittuvat ihmisten kulttuurien lukemattomat eri muodot kulttuurin kielen erityispiirteistä sosiaalisiin oloihin, hierarkioihin ja arvoihin. Tämän lisäksi, koska jo käännöstoiminta itsessään voidaan nähdä olevan kulttuurisidonnaista, onkin Tymoczkon mukaan tärkeää miettiä, mikä aika, paikka tai kieli on ollut tutkitussa prototyypissä etuoikeutetussa asemassa. (Tymoczko 1998, 654–655.) Saattaakin olla, että tästä syystä saadut tulokset eivät olekaan universaaleja, vaan muun muassa kulttuurista ja aikakaudesta riippuvaisia (Chesterman 2004, 43).

Toiseksi tutkimuksen kohteena olevien tekstien edustavuuden ongelmaksi Chesterman mainitsee sen, että jo käännöstekstejä korpukseen valittaessa tehdään päätöksiä siitä, minkälaisia käännöksiä korpukseen kelpuutetaan. Tulisiko esimerkiksi ei-ammattilaiskääntäjien tekemiä käännöksiä sisällyttää korpukseen? Entä ”huonoja” käännöksiä? (Chesterman 2004, 42.) Tämän lisäksi ongelmia käännösuniversaaliteoriassa aiheuttavat Chestermanin mukaan esimerkiksi epätarkka terminologia, jossa samoille ilmiöille on monta eri nimeä (esim. standardisaatio, yksinkertaistuminen, normalisaatio, konventionalisoituminen), eikä tutkimuksessa hänen mukaansa usein tehdä eroa lähde- ja kohdeuniversaalien välillä. Hänen mukaansa on lisäksi ongelmallista, ettei tutkimuksessa aina käy selväksi, minkälaisia käännösuniversaalihypoteeseja tutkitaan. Chestermanin mukaan onkin määriteltävä, mitä universaaleja tutkitaan: hänen mukaansa sellaisia universaaleja, jotka on havaittu kaksikielisessä rinnakkaiskorpuksesta, ei voida tutkia yksikielisestä rinnakkaiskorpuksesta (ks. luku 4.1.1), sillä kyseinen korpus ei sisällä alkuperäisiä tekstejä, joiden vertailua lähdeuniversaalien tutkiminen vaatisi. (emt, 43.) Myös käännösuniversaalien kausaalisuuden tutkiminen on ongelma, eikä käännöksissä ilmenevien kielellisten piirteiden syitä Chestermanin mukaan voida toistaiseksi kuin spekuloida (emt, 44).

Tämän lisäksi jotkin käänносuniversaaliteoriat ovat täysin vastakkaisia: miten sekä eksplikointi että yksinkertaistuminen, tai toiston poisto sekä tekstin kasvava pituus voivat olla yhtä aikaa käännosuniversaaleja? Voidaan tietenkin olettaa, etteivät kaikki universaalit esiinny samalla tavalla samoissa teksteissä ja samoissa yhteyksissä, mutta voidaanko niitä tässä tapauksessa kutsua universaaleiksi?

Näistä ongelmista huolimatta käännosuniversaalihypoteesien tutkiminen on Chestermanin mukaan ollut käännostieteelle eduksi: hänen mukaansa se on muun muassa auttanut käännostiedettä kehittämään parempia empiirisiä tutkimussuunnitelmia, tuonut korpuksat käännostieteen piiriin, ja auttanut käännostiedettä ajattelemaan yleistysten ja hypoteesien testaamisen näkökulmasta – siis auttanut kehittämään käännostiedettä empiirisenä tieteenalana. Tämän lisäksi käännosuniversaaleilla on hänestä potentiaalia myös kääntäjien koulutuksen näkökulmasta: universaaleja tutkimalla voitaisiin tunnistaa tekstiin negatiivisesti vaikuttavat universaalit, ja koulutuksen avulla auttaa opiskelijoita välttämään näitä piirteitä. (Chesterman 2010, 45). Chesterman toteaa myös, että myös kapeat yleistykset vievät eteenpäin tutkimusta ja niistä saadaan mielenkiintoisia tuloksia, vaikka ne olisikin rajattu vain tiettyyn kulttuuriin tai tekstityyppiin, ja että saattaisi kenties olla hedelmällisempää etsiä käännoissa esiintyviä, vähemmän universaaleja kaavoja, ja esittää näiden pohjalta ehkä hieman vaatimattomampia väitteitä (Chesterman 2010, 45–46). Bernardinin sekä Zanettinin mukaan näin on ollut jo pidemmän aikaa: heidän mukaansa termiä *universaali* on usein määritelty tavalla, jonka mukaan se on lähinnä yleinen tendenssi tai laaja-alainen normi (2004, 52).

3.3 Käännoisuuden aikaisempaa tutkimusta

Vaikka käännoilla onkin ollut huomattava rooli suomalaisessa kieliyhteisössä ja ne ovat vaikuttaneet suomen kirjakielen kehitykseen (Jantunen ja Eskola 2002, 184), ja jonkin aikaa suomen kielen historian aikana alkuperäissuomi ja käännoisuus ovatkin olleet lähes sama asia (Paloposki 2005, 27). Tästä huolimatta on itse käännoille tyypillisiin kielellisiin piirteisiin ollut vähäistä ennen 90-luvun loppua myös Suomessa (Mauranen ja Jantunen 2005, 1) ja vaikka tämä tutkimus on ollut pitkälti deskriptiivistä, on siihen liittynyt myös normatiivisia, käännoskieltä ja sen luettavuutta arvottavia elementtejä (esim. Vehmas-Lehto 1989; Tiina Puurtinen 1995).

Käännöskohtana käännössuomen tutkimuksessa voidaankin näin ollen pitää Savonlinnassa toiminutta Suomen Akatemian rahoittamaa tutkimushanketta, jonka seurauksena syntyneessä käännössuomen korpuksessa on yli 10 miljoonaa sanaa. Hankkeen tarkoituksena olikin kehittää käännöskielen tutkimukseen sopivia menetelmiä sekä tutkia tarkemmin juuri käännösuniversaaleja (emt., 1–2). Tästä syystä onkin tuskin ihme, että huomattava osa käännössuomen tutkimuksesta onkin 2000-luvulta eteenpäin ollut juuri korpustutkimusta. Suomessa korpustutkimusta käännössuomesta ovat tehneet muun muassa Anna Mauranen (2000, 2004), Jarmo Jantunen (2002, 2004), Sonja Tirkkonen-Condit (2004, 2005) sekä Sari Eskola (2002, 2005).

Näissä aikaisemmissa tutkimuksissa on saatu tuloksia, jotka tukevat käännösuniversaalihypoteesia. Tällaiseen loppupäätökseen on tullut esimerkiksi Sonja Tirkkonen-Condit, joka tutki liitepartikkelin *-kin* esiintymistä käännössuomessa ja alkuperäissuomessa (2005). Hän huomasi, että kyseinen liitepartikkeli, joka on englantiin ja muihin indoeurooppalaisiin kieliin suhteessa ”vastineeton”, on yleisempi alun perin suomeksi kirjoitetuissa teksteissä kuin käännössuomessa. Tämän lisäksi hän huomasi, että käännöksissä käytetyt *-kin* -ilmaukset klusteroituivat suomennoksissa alun perin suomeksi kirjoitettuja tekstejä enemmän. Käännössuomessa *-kin* esiintymät siis kasautuivat tiettyihin kaavamaisiin ilmauksiin, kun suomeksi kirjoitetuissa teksteissä kyseisen liitepartikkelin käyttö taas oli monipuolisempaa. Tirkkonen-Condit tulkitsee tuloksien tukevan käännösuniversaaliksi esitettyä hypoteesia käännöskielen kaavamaisuudesta ja kohdekielen kielellisten resurssien suppeammasta käytöstä. (Tirkkonen-Condit 2005, 133–135.) Tirkkonen-Condit on saanut samanlaisia tuloksia myös aikaisemmassa tutkimuksessaan, jossa hän tutki niin sanottujen suomen kielen ”uniikkiverbien”, kuten *ehtii*, *jaksaa* ja *riittää*, esiintymistä käännössuomessa. Hänen tutkimuksensa osoitti, että näitä verbejä esiintyi vähemmän käännössuomessa, mikä tukee Tirkkonen-Conditin mukaan käännösuniversaalihypoteesia kohdekielelle tyypillisten piirteiden aliedustumisesta käännöskielessä. (Tirkkonen-Condit 2004, 179–181.)

Myös Sampo Nevalainen, joka tutki suomennosten sanastoa taajuuslistojen avulla, on saanut omassa tutkimuksessaan samankaltaisia tuloksia. Nevalaisen tutkimuksen mukaan sana-sane -suhde on pienempi suomennoksissa, ja suomennoksissa yleisempien sanojen osuus kaikista sanoista on prosentuaalisesti korkeampi (Nevalainen 2005, 146–147). Hän huomasi myös sanaston olevan suomennoksissa yhtenäisempi: usein käytetyt sanat ylikorostuvat, ja harvoin käytettyjä sanoja esiintyy vähemmän kuin alkuperäissuomessa (emt., 135). Näin ollen myös Nevalaisen tulokset näyttävät tukevan yksinkertaistumishypoteesia.

Kaikki aikaisempi tutkimus ei kuitenkaan ole tukenut käännösuniversaalihypoteeseja. Esimerkiksi Jarmo Jantunen tuli väitöskirjassaan siihen tulokseen, että vaikka todisteita epätyypillisistä frekvensseistä löytyi (2004, 215), ei todisteita käännöskielen yksinkertaistumisesta ollut selvästi havaittavissa (emt. 218). Sen sijaan hän panee merkille, että lähdekielellä näytti olevan vaikutus tutkimuksessa saatuihin tuloksiin (emt. 219). Lastenkirjallisuuden suomennosten eksplisiittistymistä konnektiivien kautta tutkinut Tiina Puurtinen (2004) taas tuli tutkimuksessaan siihen lopputulokseen, että vaikka jotkut konnektiivit esiintyivätkin käännössuomessa alun perin suomeksi kirjoitettuja tekstejä enemmän, käyttäytyivät jotkut muut konnektiivit päinvastoin. Näin ollen Puurtisen tulokset eivät hänen oman luonnehdintansa mukaan liioin tukeneet tai olleet ristiriidassa eksplisiittistymishypoteesin kanssa. (Puurtinen 2004, 174). Kaiken kaikkiaan käännösuniversaaleja tutkivien tutkimusten tulokset ovat siis olleet jokseenkin vaihtelevia: vaikka joissakin tutkimuksissa käännösuniversaaleihin viittaavia piirteitä löytyy (esim. Mauranen 2005, Eskola 2005, Nevalainen 2004), on muista tutkimuksista saatu päinvastaisia (Jantunen), tai ainakin vähemmän selkeitä (Puurtinen 2004).

Lisäksi käännösuniversaalien tutkimus on rajoittunut pitkälti journalistisiin ja kaunokirjallisiin teksteihin, eikä käännösuniversaaleja ole tutkittu juurikaan virkakielen tai teknisen kielen osalta. Poikkeuksena tästä on Johanna Pullan pro gradu -tutkielma (2011), joka vertaili temporaalirakenteita Euroopan keskuspankin *Kuukausikatsauksen* ja Suomen Pankin *Euro & Talous* sekä *Talouden näkymät* teksteistä koostuvien korpusten välillä. Tämän lisäksi Svetlana Probirskaja on tutkinut väitöskirjassaan Suomen ja Venäjän välisiä sopimuksia korpusmenetelmin, joskaan hän ei analysoinut korpuksia käännösuniversaalien näkökulmasta (2009). Näin ollen käännösuniversaaleihin keskittyvä korpustutkimus laki- ja virkakielen tekstilajista tuo käännösuniversaaliteoriaan uuden näkökulman, jonka lisäksi on mahdollista saada uutta tietoa lain ja hallinnon käännösteksteistä itsestään.

4. KORPUSLINGVISTIIKKA

Tässä luvussa esittelen korpustutkimuksen luonnetta ja määritelmiä pääpiirteittäin. Lisäksi luvussa pohditaan korpustutkimuksen luonnetta kvantitatiivisena menetelmänä ja tarkastellaan sen etuja ja heikkouksia käännöstieteellisenä tutkimusmenetelmänä.

4.1 Korpuksen määritelmä ja korpuksen kokoaminen

Vaikka sana *korpus* onkin alun perin tarkoittanut erityisesti saman kirjoittajan kirjoituksista koostuvaa tekstikokoelmaa, viittaa termi Mona Bakerin mukaan käännöstieteessä nykyään pääosin elektronisessa muodossa olevaan joukkoon tekstejä, jotka on koottu tiettyä tarkoitusta varten (Baker 1995, 225). Korpus voidaan myös mieltää tietyn periaatteen mukaan valikoiduksi suureksi joukoksi luonnollisia tekstejä (Biber ym. 1998, 12) tai rajalliseksi kokoelmaksi kieliopillisia lauseita, jota käytetään kielen deskriptiivisen analyysin pohjana (Newton 1993, 223, ref. Baker 1995, 225). Termiä on kuitenkin myös käännöstieteen piirissä käytetty suppeasta joukosta tekstejä, joita analysoidaan manuaalisesti (Baker 1993, 241).

Korpus ei kuitenkaan ole vain joukko tekstejä, vaan sen tehtävänä on kuvata kieltä tai jotain osaa kielestä. Tästä puolestaan seuraa, että korpuksen rakenne riippuu siitä, mitä kieltä tai osaa kielestä korpus pyrkii kuvaamaan. Tämä taas vaikuttaa siihen, minkälaisia tutkimuskysymyksiä korpuksella voidaan tutkia ja kuinka yleistettävissä saadut vastaukset ovat. (Biber ym. 1998, 246). Korpuksen valittujen tekstien tulisikin noudattaa selviä valintakriteerejä, jotta voidaan varmistaa, että se on edustava näyte kielen alueesta tai tekstityypistä, jota se käsittelee (Baker 1995, 225). Näitä valintakriteerejä ovat muun muassa kielen tyyppi (puhuttu vs. kirjoitettu kieli), alueelliset rajat (amerikanenglanti vs. brittienglanti) ja kieli itse (yleiskieli vs. akateeminen kieli) (Baker 1995, 229). Baker kuitenkin huomauttaa, että nämä kriteerit ovat käännöstieteellisen tutkimuksen tarkoitukseen vajaat, ja ne ovat lingvistien kehittämiä, yksikielisiä korpuksia koskevia kriteereitä (Baker 1995, 229–230).

Douglas Biber, Susan Conrad ja Randi Reppen taas nostavat esille tilastotieteellisiä kriteerejä, joiden tarkoituksena on varmistaa tekstien oikea representatiivisuus. Nämä kriteerit ovat otoksen valinta, otoksen monipuolisuus ja korpuksen koko (Biber ym. 1998, 246–248). Otoksen valinta määrittää sen, miten ja minkälaisia tekstejä korpukseen valitaan, esimerkiksi valitaanko tekstit suhteellisella otannalla (*representative sample*) vai ositetulla otannalla (*stratified sample*) (Biber ym. 1998, 247–248). Biber ja hänen kollegansa kuitenkin huomauttavat, että kyseiset määritelmät ovat kielitieteen kannalta jokseenkin ongelmallisia, sillä esimerkiksi suhteellinen otanta kielestä ei välttämättä kattaisi rekistereitä, kuten laki- tai lääketieteellistä kieltä, eikä se näin ollen olisi edustava otos kielestä (emt. 247). Heidän mukaansa korpuksen suunnittelun tärkein osa useimmissa tutkimuksissa onkin tunnistaa kielen kielelliset variaatiot ja huolehtia siitä, että ne tulevat edustetuksi, ja heistä korpustutkimuksessa on tärkeää ottaa huomioon, miten eri kategoriat ja alakategoriat on identifioitu, ja mitä kriteereitä korpuksen kategorisointiin on käytetty (emt. 247–248).

Myös valittujen tekstien monipuolisuus on tärkeä osa edustavaa korpusta luotaessa. Monipuolisuuden eri aspekteiksi Biber ja kollegat luettelevat muun muassa eri rekisterit, aksentit ja murteet sekä tekstien aihepiirit. Monipuolisuus on erityisen tärkeää leksikografisissa tutkimuksissa siksi, että eri sanat esiintyvät eri aihepiireissä eri tiheyksillä. Lisäksi korpuksen koko on edustavuuden kannalta tärkeää: liian suppeassa korpuksessa yksi teksti voi vaikuttaa liikaa analyysin tuloksiin. (Biber ym. 1998, 248–249). Lisäksi tekijänoikeuskysymykset saattavat joissain tapauksissa nousta ongelmaksi korpusta koottaessa (Baker 1995, 234–235; Olohan 2004, 50–51).

Kaiken kaikkiaan edustavan korpuksen tulee siis olla mahdollisimman laaja ja kuvata kielen eri aspekteja mahdollisimman monipuolisesti. Kielen monipuolisuutta koskeva kriteeri osoittautuu kuitenkin jokseenkin ongelmalliseksi tutkittaessa erikoiskieliä, kuten instituution tuottamia tekstejä tai lain ja hallinnon tekstejä yleensä. Se, mitä ”monipuolisuus” tässä yhteydessä tarkoittaa, jää siis näiltä osin tutkijan määriteltäväksi, ja tästä enemmän luvussa 5.

4.1.1 Erilaisia korpustyypppejä

Korpukset on mahdollista luokitella niissä olevien tekstien ja kielten mukaan useisiin eri kategorioihin. Korpukset voidaan esimerkiksi luokitella korpukseen kerätyn materiaalin tyypin (puhekielen korpus, kirjoitetusta kielestä koostuva tekstikorpus), siinä käsitellyn materiaalin kielen (yksikielinen, kaksikielinen tai monikielinen korpus) tai korpuksessa tutkittavien tekstien (rinnakkaiskorpus, verrannollinen korpus) mukaan. Koska kaksi ensimmäistä luokittelukategoriaa ovat varsin selkeitä konsepteja, keskityn tarkemmin selventämään vain rinnakkaiskorpuksen ja verrannollisen korpuksen käsitteitä.

Rinnakkaiskorpus (*parallel corpus*) sisältää alkuperäiset, lähdekieliset tekstit sekä niiden käännökset yhdellä tai useammalla kielellä. On huomautettu, että termin *rinnakkaiskorpus* käyttö tässä yhteydessä ei ole täysin ongelmaton, sillä samalla termillä voitaisiin viitata myös kaksikielistä rinnastettavista teksteistä koostuvaan korpukseen, joka ei kuitenkaan koostu käännöksistä, ja tästä syystä onkin tilalle ehdotettu myös termiä *käännöskorpus* (Olohan 2004, 24). Käännöstieteessä rinnakkaiskorpuksen tekstit usein *kohdistetaan* (*align*), joka tarkoittaa sitä, että tekstin kielelliset yksiköt kohdistetaan vastaaviin yksikköihin korpuksen toisessa tai useammassa kielessä (Olohan 2004, 26). Onkin ehdotettu, että rinnakkaiskorpuksia voitaisiin käännöstieteessä näin ollen käyttää vaihtoehtona ääneenajattelumetodille kääntämisen prosessia ja käännöksen tuloksia tutkittaessa yhdistämällä korpustutkimus ja kääntäjiltä kerätty introspektiivinen data, kuten päiväkirjat. Lisäksi on esitetty, että itse käännösratkaisuja tutkimalla pystytään tutkimaan käännösprosessia, ja että korpuksella voidaan tunnistaa ja mitata näitä piirteitä. (Emt., 38–39.)

Verrannollinen korpus (*comparable corpora*) puolestaan on yksikielisestä tekstmateriaalista koostuva korpus, jonka käytöstä käännöstieteellisessä tutkimuksessa esimerkiksi Mona Baker on kirjoittanut runsaasti (ks. esim. Baker 1995). Yleisin tapa käyttää verrannollista korpusta käännöstieteen tutkimuksessa on koostaa kaksi korpusta, joista toinen koostuu alkuperäiskielellä laaditusta materiaalista, ja toinen kyseiselle kielelle käännetystä materiaalista. Verrannollinen korpus onkin juuri se metodi, jota käyttämällä monia käännösuniversaalihypoteeseja on tutkittu (ks. luku 3).

4.1.2 Korpus ja käännöstutkimus

Korpuksen käytöllä osana käännöstieteen tutkimusta on hyvin lyhyt historia verrattuna korpuksen käyttöön osana lingvistiikan tutkimusta (Olohan 2004, 13), mutta korpustutkimus on kasvattanut suosiota käännöstieteen piirissä 1990-luvulta lähtien.

Artikkelissaan *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research* (1995) Mona Baker esittelee korpukseen pohjautuvaa käännöstieteen tutkimusta sekä sen tarjoamia uusia tutkimusmahdollisuuksia. Hän mainitsee esimerkkinä muun muassa rinnakkaiskorpusten käytön tietokonepohjaisessa käännösopetuksessa ja mainitsee, että niiden avulla voidaan saada objektiivisia esimerkkejä siitä, miten kääntäjät ratkaisevat käännösongelmat käännöksissään (emt., 231).

Samassa artikkelissa Baker käsittelee myös verrannollista korpusta ja kuinka sitä olisi mahdollista hyödyntää käännösteksteissä alkuperäiskieltä useammin esiintyvien tai käännöskielelle ominaisten elementtien tutkimisessa (1995, 234). Näistä elementeistä hän mainitsee erikseen sana-sane -suhteen, jonka tutkiminen käännetyissä teksteissä voi Bakerin mukaan auttaa tunnistamaan globaaleja piirteitä, jotka identifioivat käännökset käännöksiksi. Bakerin mukaan sana-sane -suhde saattaa myös kertoa jotain käännöksissä tapahtuvan mediaation luonteesta. Samassa artikkelissa Baker ehdottaa myös käännösten ja lähdekielisten tekstien leksikaalisen tiheyden tutkimista. (Emt., 236–238).

Korpustutkimus onkin kasvattanut suosiotaan käännöstieteessä 90-luvulta eteenpäin, ja sitä onkin käytetty runsaasti erityisesti käännösuniversaalien tutkimiseen (ks. luku 3.3), mutta sitä on käytetty myös muussa käännöstieteen tutkimuksessa. Yksi korpustutkimuksen muoto käännöstieteessä on kääntämisen tyylin ja normien tutkiminen. Tämän lisäksi korpusmenetelmillä on tutkittu myös esimerkiksi tulkkausta ja puhutulle kielelle tyypillisiä piirteitä sekä kielenoppimista. (Zanettin 2013, 21).

4.2 Korpustutkimus menetelmänä

Douglas Biberin, Susan Conradin ja Randi Reppen mukaan kielentutkimus voidaan jakaa kahteen pääkategoriaan: kielen rakenteen tutkimukseen ja kielen käytön tutkimukseen. Siinä, missä kielen rakenteen tutkimus tutkii kielen piirteitä ja rakenteita ja sitä, miten kielen pienemmät yksiköt muodostavat keskenään suurempia kokonaisuuksia, tutkii kielen käytön tutkimus esimerkiksi sitä, miten puhuttu ja kirjoitettu kieli eroavat toisistaan, tai miten kieli eroaa eri käyttötilanteissa ja rekistereissä. Tätä tutkiakseen tutkijat pyrkivät analysoimaan yleisiä ja toistuvia säännönmukaisuuksia sekä kontekstuaalisia tekijöitä, mikä on haaste metodologian kannalta. Manuaalinen analyysi on ongelmallinen siksi, että Biberin ja hänen kollegoidensa mukaan ihmiset huomaavat helpommin epätyypilliset kuin tyypilliset piirteet. Lisäksi tutkimusmateriaalin kerääminen ja läpikäyminen muodostuisi ongelmaksi. Korpustutkimus tarjoaa kuitenkin mahdollisuuden käsitellä suurta määrää kielellistä dataa ja huomioida kontekstuaaliset tekijät. (Biber ym. 1998, 1–3).

Biber ja kollegat määrittelevät korpus-pohjaisen lähestymistavan tutkimustavaksi, jonka piirteitä ovat empiirinen tutkimustapa, jossa analysoidaan luonnollisissa teksteissä esiintyviä tyypillisiä piirteitä käyttämällä korpusta analyysin pohjana, ja jossa käytetään hyväksi tietokoneanalyysin sekä automatisoituja että interaktiivisia tekniikoita, ja tulosten analysoimisessa käytetään sekä kvantitatiivisia että kvantitatiivisia analyysitekniikoita. Tästä syystä korpuksen käyttö tutkimusmetodina mahdollistaa entistä laajempien ja tarkempien analyysien tekemisen. (Biber ym. 1998, 4). Lisäksi korpus mahdollistaa kieliaineksen laajan rakenteellisen tutkimisen: esimerkiksi sanaluokkajäsennetty korpus mahdollistaa kieliopin kvantitatiivisen tutkimuksen tavalla, johon ei ole ennen korpuksen käyttöä ole ollut mahdollisuutta.

On kuitenkin muistettava, että kielen, tai edes vain kielen yhden osan tai aspektin kuvaaminen on ongelmallinen tehtävä, sillä kielen kaikki muotoja tai kontekstuaalisia muuttujia on mahdoton määritellä, ja näin ollen on mahdotonta antaa kattavaa kuvaa kielen koko kirjosta (Biber ym. 1998, 246). Tämän lisäksi on myös otettava huomioon, että korpuksen koolla on suuri vaikutus siihen, millaisia tuloksia korpuksella saadaan. Liian pienessä korpuksessa yhdellä tekstillä saattaakin olla liian suuri vaikutus analyysin tulokseen (emt., 249). Lisäksi korpustutkimuksen tuloksia eritellessä on otettava huomioon, että korpus antaa raakaa dataa, jota on pystyttävä tulkitsemaan mielekkäällä

tavalla. Tulosten on siis oltava vertailtavissa ja ne on tarpeen mukaan normalisoitava, jotta tämä on mahdollista (emt. 263). Sanaluokkajäsennettyä korpusta käytettäessä on myös otettava huomioon, että jäsennysprosessi on tietokonepohjainen, eikä se aina onnistu tunnistamaan lauseenjäseneitä oikein. Tämän vuoksi lausejäsennetty korpus ei ikinä voi olla täysin tarkka, vaikka monet tutkimukset jäsennysohjelmien tarkkuudesta raportoivatkin tarkkuuden olevan yli 90 prosentin luokkaa (Biber ym. 1998, 262).

Kaiken kaikkiaan korpuksset kuitenkin tarjoavat kvantitatiivisen, empiirisen menetelmän käännösten tutkimiseen ja niiden leksikaalisten sekä kielen rakenteellisten piirteiden laajempaan tarkasteluun, joka on tutkimuskohteena juuri käännösuniversaalihypoteeseissa. Näin ollen, vaikka korpusmenetelmien käyttöön liittyy ongelmia, on se hyödyllinen tutkimusmenetelmä, jolla voidaan oikein käytettynä saada tuloksia, joita voidaan yleistää kvantitatiivisten menetelmien ongelmista huolimatta huomattavasti laajemmin kuin esimerkiksi suppeammasta kvalitatiivisesta tutkimuksesta saatuja tuloksia. Korpustutkimuksessa tarkkuus siis vaihdetaan osittain volyyymiin, mikä mahdollistaa uusien näkökulmien tutkimisen, mutta joka pitää myös ottaa huomioon tutkimustuloksia tarkasteltaessa.

4.3 Korpustutkimuksen metodit käännösten tutkimuksessa

Koska korpus itsessään on vain elektroninen joukko tekstejä, on sen sisältämää dataa käsiteltävä ja jäsenneltävä, jotta sitä voidaan käyttää tutkimuksessa. Korpusohjelmissa onkin useita eri sovelluksia ja tapoja tulkita korpuksen sisältämää kieliaainesta. Näitä sovelluksia käyttämällä voidaankin tutkia useita kielen eri piirteitä.

Yksinkertaisimpia korpussovelluksia korpuksessa olevien yksiköiden (sanojen, fraasien, päätteiden) tutkimiseen ovat *frekvenssi* sekä *konkordanssi*: yksikön frekvenssi kertoo, kuinka monta kertaa se esiintyy korpuksessa, kun taas konkordanssi ilmaisee laajemman yhteyden, jossa sana esiintyy. Frekvenssejä käyttämällä voidaan korpuksen avulla laatia niin sanottu frekvenssilista (*frequency list*), jota kutsutaan myös sanalistaksi. Tämä listaa jokaisen korpuksessa käytetyn sanan ja sen, kuinka monta kertaa se korpuksessa esiintyy. Tätä työkalua käyttämällä voidaan tunnistaa korpuksen

yleisimmät sanat, ja tätä kautta puolestaan kartoittaa yleisimmät lekseemit tai epätavalliset kielelliset ilmentymät (Olohan 2004, 77).

Jotkin korpuksset, kuten WordSmith, laskevat myös niin sanotun *sana-sane -suhteen* (*type-token ratio*), jossa korpusohjelma laskee korpuksessa esiintyvien yksittäisten sanojen ja kaikkien korpuksessa olevien sanojen suhteen: sitä voidaan siis käyttää mittaamaan sitä, kuinka monipuolista korpuksen sanasto on. Mikäli korpus A:n sana-sane -suhde on pienempi kuin korpus B:n, on korpus A:n teksteissä siis enemmän toistoa (Olohan 2004, 81), ja näin ollen myös sen sanasto on yksinkertaisempaa. Sana-sane -suhteiden vertailu onkin yksi mahdollinen tapa tutkia käännösuniversaaliteorian yksinkertaistumishypoteesia (esim. Nevalainen 2005, ks. luku 3.3). WordSmith laskee myös niin sanotun *standardisoidun sana-sane -suhteen* (*standardized type-token ratio*, *STTR*), joka laskee sana-sane -suhteen aina 1000 sanan välein. Tällä tavalla voidaan luotettavammin vertailla sellaisia korpuksia, jotka koostuvat eripituisista teksteistä (Olohan 2004, 81).

Konkordanssia apuna käyttämällä taas voidaan tutkia sitä, missä yhteydessä eri sanat korpuksessa esiintyvät, eli millaisia kollokaatteja niiden yhteydessä esiintyy: korpuksen konkordanssityökalujen käyttämisestä tällä tavalla kutsutaan *avainsana-analyysiksi* (eng. *key word in context*, *KWIC*) (Olohan 2004, 63). Tällä tavalla voidaan puolestaan tutkia käännösuniversaaliteorian hypoteesia käännöskielen epätyypillisistä frekvensseistä (esim. Mauranen 2000, Jantunen 2004). Tämän lisäksi konkordanssia voidaan käyttää määrittämään esimerkiksi tietyn fraasin tai päätteen ilmenemisen määrä korpuksessa: frekvenssin avulla selviää vain, kuinka monta kertaa tietty sana ilmenee korpuksessa. Konkordanssia käyttämällä sen sijaan voidaan erilaisia hakuehtoja käyttämällä tutkia esimerkiksi sitä, minkälaisia kielellisiä elementtejä rinnastetaan konjunktiolla *ja*, ja verrata näitä ilmaisulla *sekä – että* tehtyihin rinnastuksiin.

Näiden ominaisuuksien lisäksi käsitellyillä korpuksilla voidaan tutkia korpuksen tekstien ominaisuuksia vielä tarkemmin. Lemmatisoitu korpus lukee kaikki yhden sanan samat muodot (olen, olemme, ollaan) samaksi lemmaksi eli sanan perusmuodoksi (olla). Tämän avulla päästään tarkastelemaan tekstissä esiintyviä yksittäisiä sanoja, jonka kautta voidaan tutkia luotettavammin tekstissä käytetyn sanaston monipuolisuutta vertailemalla esimerkiksi korpuksessa olevien uniikkien lemموjen määrää, tai korpuksen yleisimpien lemموjen prosentuaalista suhdetta kaikkiin korpuksen

lemmoihin, jonka kautta voidaan myös tutkia käännösuniversaaliteorian yksinkertaistumishypoteesia (esim. Nevalainen 2004).

Lausejäsenneyllä korpuksella taas päästään tutkimaan korpuksen leksikaalisia ominaisuuksia, eli sitä, minkälaisista sanoista teksti koostuu: kuinka paljon siinä on verbejä, substantiiveja, ja niin edelleen. Tämä mahdollistaa kielen rakenteellisen vertailun ja muun muassa leksikaalisen tiheyden laskemisen. Leksikaalinen tiheys ilmaisee niin sanottujen leksikaalisten sanojen (sisältösanojen) suhteen kaikkiin korpuksen sanoihin verrattuna (Olohan 2004, 200). Tämän lisäksi korpuksella voidaan tarkastella tekstien muita kielellisiä piirteitä, kuten sanojen tai lauseiden keskimääräisiä pituuksia.

Korpuksot tarjoavat siis loistavan mahdollisuuden tarkastella suurta joukkoa tekstejä kvantitatiivisia menetelmiä käyttäen. Tämä puolestaan mahdollistaa käännöstekstien ja niissä käytetyn kielen tarkastelemisen uudesta näkökulmasta: niiden avulla voidaan vertailla systemaattisesti tekstien eri piirteitä. Tästä syystä ne ovatkin hyvä tutkimuskeino, kun on tarve saada nopeasti ja helposti kielellistä dataa laajoista tutkimusaineistoista, ja tästä syystä ne sopivat varsinkin suurempien tekstikokonaisuuksien, kuten tietyille tekstilajille tyypillisten piirteiden, tarkasteluun. Tämän lisäksi niiden käyttö mahdollistaa sellaisen aineistovolyymien, jonka pohjalta voidaan tehdä päätelmiä käännösuniversaalihypoteesien tapaisista laaja-alaisista konsepteista. Juuri näistä syistä korpustutkimus sopiikin sekä valitsemaani tutkimusnäkökulman käsittelyyn että tutkimuskohteekseni valitsemieni tekstien tutkimiseen.

5. TUTKIMUSMATERIAALI JA METODIT

Luvussa esittelen tutkimusaineistoni ja niiden valitsemisprosessin sekä tutkimuksessa käyttämäni korpuksset. Tämän lisäksi selostan, mitä tutkimusmetodeita ja korpusmenetelmiä tutkimuksessa käytän ja selitän, miten niitä voidaan käyttää juuri käänösuniversaalihypoteesien tutkimiseen.

5.1 Aineisto

Koska tutkimukseni kohteena on nimenomaan lakikieli, olen valinnut tutkimusaineistokseni lainvoimaisia tekstejä, jotka ovat juridisesti sitovia, jotta käyttämäni aineistot olisivat mahdollisimman vertailukelpoisia. Tästä syystä suomennoksista koostuvaan aineistooni on valikoitunut vain Euroopan unionin asetuksia: ne ovat Euroopan unionin teksteistä ainoita, jotka ovat kaikkia jäsenmaita sellaisenaan sitovia. Suomenkielinen aineistoni puolestaan koostuu Suomen säädöskokoelmasta poimituista säädöksistä. Olen kummassakin tapauksessa pyrkinyt pitämään aineiston mahdollisimman tuoreena, enkä ole valinnut Suomen säädöskokoelmasta säädöksiä, jotka on kirjoitettu ennen vuotta 1995: suurin osa materiaalista onkin kirjoitettu 2000-luvun puolella. Näitä periaatteita lukuun ottamatta olen valikoinut korpuksiin tekstit satunnaisotantana: ainoana poikkeuksena on, että olen parhaani mukaan pyrkinyt varmistamaan, että kummastakin korpuksesta löytyy ainakin jossain määrin saman aihepiirin testejä. Kaikki tekstit ovat julkisesti saatavilla Oikeusministeriön FinLex -säädspalvelusta sekä Euroopan unionin EUR-Lex -palvelusta. Koska viranomaisten ja muiden julkisten elinten päätökset eivät saa tekijänoikeussuojaa, on kyseisten tekstien kopioiminen ja käyttäminen korpusmateriaalina luvallista.

Euroopan unionin suomenkielisiä asetuksia sisältävään korpukseen on valittu asetuksia lähes kaikilta mahdollisilta oikeuden aloilta siviili- ja rikosoikeudesta ympäristöä koskevaan lainsäädäntöön (Liite B). Tällä pyritään minimoimaan yksittäisessä aihepiirissä mahdollisesti esiintyvän kielellisen piirteen vaikutus. Sama pätee Suomen lainsäädännön teksteihin (Liite A), joiden valinnassa on myös pyritty siihen, ettei mukaan ole otettu useaan otteeseen suoraan Euroopan unioniin vastaaviin asetuksiin, direktiiveihin tai suosituksiin viittaavia lakeja, jotta mahdollisilta päällekkäisyyksiltä korpuksien välillä välttyttäisiin.

Korpuksessa käytetyistä teksteistä on poistettu formatiiviset alkutekstit, jotka toistuvat jokaisessa Euroopan unionin asetuksessa ilmaisemaan, ketä ja mitä tahoja on lain valmisteluprosessin aikana kuunneltu, jotta juuri näissä yleisesti toistuvat lekseemit ja fraasit (*komissio, parlamentti, ottaa huomioon, katsoo seuraavaa*) eivät vääristäisi tuloksia. Asetusten voimaantuloa koskeva artikla sekä allekirjoitus on jätetty korpusmateriaalista pois samasta syystä. Näin on tehty myös suomalaisten lakien kohdalla: lakien alkufraasi ”*Eduskunnan päätöksen mukaisesti säädetään*”, voimaantuloa koskeva lainkohta sekä muutossäädösten voimaantuloa ja soveltamista koskevat maininnat on jätetty aineistosta pois samoista edellä mainituista syistä.

Korpukseseen ei ole myöskään sisällytetty asetuksien tai säädösten yhteyteen lisättyjä liitteitä, sillä ne sisältävät usein esimerkiksi kaavioita ja käsitteenmäärittelyjä, jotka eivät aina sisällä kokonaisia lauseita, ja joiden tutkiminen korpusaineistossa ei tästä syystä ole tämän tutkimuskysymyksen puitteissa mielekäästä. Lisäksi suomenkielisistä teksteistä on poistettu pykälää tarkoittava merkki § tiedonkäsittelyyn liittyvien ongelmien ehkäisemiseksi, kuten myös irralliset päätteet (kuten ”:n” sekä ”:ssä”) jotka korpus lukee sanoiksi ja jotka näin ollen vääristävät analyysin tuloksia.

5.2 Korpuksat

Euroopan unionin asetuksien käännöksiä sisältävä korpus (EU-korpus) koostuu 24 tekstistä, ja se on kooltaan hieman yli kaksisataatuhatta sanaa (202 919). Suomen säädöskokoelmasta poimituista, alkuperäiskieleltään suomenkielisistä teksteistä koostuva korpus (SU-korpus) on kooltaan hieman pienempi, vaikka se sisältääkin kappalemäärältä enemmän tekstejä (36 kappaletta, Liite 2), mutta sekin ylittää yli kahdensadantuhannen sanan (200 680). Näin ollen korpuksat ovat suurin piirtein vertailukelpoiset, sillä niiden välinen kokoero on vain 3,6 prosenttia, vaikka erot on tietysti otettava huomioon tuloksia tarkastellessa, minkä lisäksi myös tekstien eri lukumäärä voi aiheuttaa ongelmia aineistoa analysoitaessa. Kyseessä on tietenkin erimerkiksi käännössuomen korpuksen 11 miljoonaan sanaan verrattuna pieni aineisto, mutta koska kyseessä on erikoiskielen korpus joka koostuu kapea-alaisesta joukosta tekstejä, voidaan myös kooltaan pienempää korpusta käyttää tutkimusvälineenä (Zanettin 2011, 15). Lisäksi Douglas Biberin tutkimuksen mukaan kymmenen eri tekstin otos tekstilajista edustavat hyvin useita kieliopillisia piirteitä (Biber 1990, ref. Biber ym. 1998, 249). Koska molemmissa korpuksissa on tätä huomattavasti enemmän yksittäisiä tekstejä, pitäisi korpuksen ainakin periaatteessa antaa hyvä läpileikkaus lain ja hallinnon tekstien piirteistä. Tästä huolimatta

korpuksen suhteellisen pieni koko tulee kuitenkin pitää mielessä, kun saatuja tutkimustuloksia tulkitaan.

On myös hyvä ottaa huomioon, että sanamäärät ovat vain suuntaa-antavia, ja että lakitekstejä korpuksen avulla analysoitaessa korpusohjelma tulkitsee juoksevassa tekstissä esimerkiksi numeeriset viittaukset muihin lakeihin ja lainkohtiin sekä niihin liittyvät lyhenteet ja merkinnät, kuten (395/1987), sanoiksi. Valitettavasti aineiston editointi kuvastamaan tätä olisi niin työlästä, ettei se ole käytännössä mahdollista. Tämä on kuitenkin lakikielelle tyypillinen piirre, ja sitä esiintyy molemmissa korpuksissa, joten sen ei pitäisi vaikuttaa tuloksiin tekstien kielellisiä piirteitä vertaileissa.

5.3 Tutkimuskysymys ja -metodit

Keskityn tutkimaan universaalihypoteeseista erityisesti yksinkertaistumista, jota tutkin Nevalaisen (2005) tapaan tutkimalla korpuksen sana-sane -suhteita. Tämän lisäksi tutkin yleisemmällä tasolla sanojen ja virkkeiden keskimääräisiä pituuksia, jossa vertailukohtanani on Vesa Heikkisen, Outi Lehtisen sekä Mikko Lounelan suomenkielisiä tekstejä 90-luvulta sisältämästä Parole-korpuksesta tekemä analyysi (2000). Korpusohjelmana käytän puolestaan Mike Scottin kehittämää WordSmith Toolsin versiota 6, ja aineistooni on tehty sanaluokka-annotaatio käyttämällä Helmut Schmidin kehittämää TreeTagger -ohjelmaa. Valitettavasti suomen kielen lemmatisaatio on suomen kielen ja sanojen rakenteeseen nähden jokseenkin ongelmallista eikä kovin tarkkaa, ja se on tästä syystä jätetty pois tämän tutkimuksen piiristä.

Tutkin myös korpuksien leksikaalista tiheyttä ja sanaluokkajakaumaa ja pyrin tätä kautta selvittämään, onko niissä huomattavia eroja ja ovatko luvut samansuuntaisia Pirjo Hiidenmaan (2000, 44) saamien tuloksien kanssa. Tätä kautta haluan tutkia yli- tai alikorostuvatko jotkin näistä virka- ja lakikielelle tyypillisistä piirteistä, kuten korkea leksikaalinen tiheys ja substantiivivoittoisuus, jommassakummassa korpuksessa. Laviosan saamien tulosten mukaan leksikaalinen tiheys on käännöksissä alkuperäiskieltä matalampi (Laviosa 1998, 557), mikä viittaisi puolestaan siihen, että käännöskieli yksinkertaistuu sanastollisen tason lisäksi myös rakenteellisella tasolla.

Aion siis näitä parametreja käyttämällä tutkia yksinkertaistuuko lakikielen käännössuomi sana- tai rakennetasolla. Analysoin tätä vertailemalla eri korpuksien sana-sane -suhteita sekä niiden leksikaalista tiheyttä. Tämän lisäksi tarkastelen korpuksien tekstien muita kielellisiä ominaisuuksia, kuten lausepituuksia, sanojen keskimääräisiä pituuksia ja sanaluokkakajakaumia ja vertailen näitä keskenään ja selvitän, onko näissä huomattavia eroja ja vastaavatko ne aikaisempaa lain ja hallinnon teksteistä toteutettujen tutkimusten tuloksia. Aikaisemman tutkimuksen pohjalta hypoteesina on, että sana-sane -suhde on EU-tekstejä sisältävässä korpuksessa pienempi, ja sen leksikaalinen tiheys matalampi.

6. ANALYYSI

Tässä luvussa keskityn erittelemään korpuksesta saatua dataa ja käsittelemään sitä luvussa 4 eriteltyjen korpusmetodien avulla.

6.1 Yleistä korpuksista: sana- ja virkepituuksia

Aluksi vertailen eri korpuksien tilastollisia lukuja virkkeiden ja sanojen pituuksien osalta ja vertailen saatuja tuloksia Hiidenmaan (2000) sekä Heikkisen ym. (2001) Oulun korpuksen pohjalta tekemään vastaavaan analyysiin.

6.1.1 EU-korpus

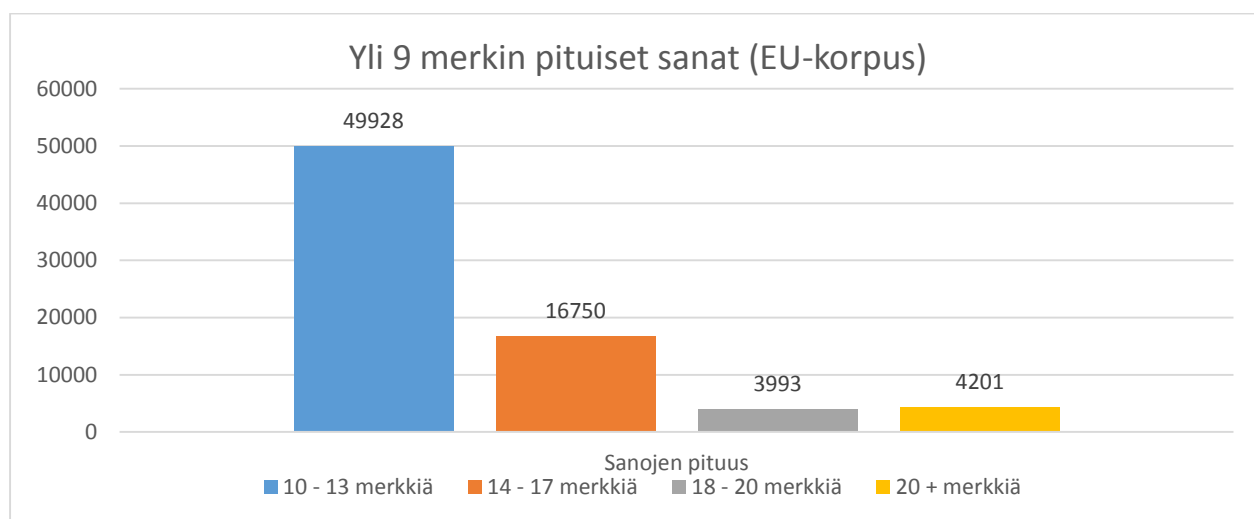
Kuten jo ylempänä todettiin, Euroopan unionin teksteistä koostuva korpus koostuu 202 919 sanasta, joista on erotettavissa 20 215 erillistä sanaa. Korpuksessa on hieman yli 6 500 virkettä, joiden sanamäärä ja sanojen pituus ovat seuraavanlaiset:

Taulukko 1: Sana- ja virketilastoja (EU-korpus)

sanojen keskim. pituus (merkeissä)	sanojen pituus (keskihajonta) (merkeissä)	virkkeiden määrä	virkkeen keskim. pituus (sanoissa)	virkkeen keskim. pituus (keskihajonta) (sanoissa)
8,91	4,96	6645	28,51	34,60

Prosentuaalinen ero sanojen keskimääräisessä merkkimäärän ja keskihajonnan mukaan lasketun merkkimäärän välillä on yli 44 prosenttia, mikä on jokseenkin huomattava ero. Tämä ero sanojen pituuden keskiarvon ja keskihajonnan välillä viittaisi siihen, että sanojen pituus teksteissä vaihtelee jokseenkin paljon: sekä pitkiä että lyhyitä sanoja on siis runsaasti. Myös virkkeiden pituus vaihtelee: prosentuaalinen ero keskimääräisen pituuden ja keskihajonnan mukaan lasketun pituuden välillä on noin 17,6 prosenttia. Tämä prosentuaalinen ero on kuitenkin huomattavasti pienempi kuin sanojen keskimääräisen pituuden ja niiden pituuden keskihajonnan välinen vaihtelu. Siis, sanojen merkkipituudet saattavat tekstissä vaihdella huomattavastikin, mutta virkkeiden pituus sanoissa laskettuna pysyy jokseenkin samana, vaikka vaihtelua onkin runsaasti.

Lukujen vertailukohtana voidaan käyttää esimerkiksi 90-luvulta peräisin olevia suomenkielisiä tekstejä sisältävää Parole-korpusta, sekä 60-luvun suomea sisältävää Oulun korpusta, joita Vesa Heikkinen, Outi Lehtinen sekä Mikko Lounela tutkivat artikkelissaan *Kuvia kirjoitetusta suomesta* (2001). Vertailemalla saatuja lukuja Heikkisen ym. löytöihin voidaan todeta, että korpuksen virkkeet ovat huomattavasti pidempiä kuin yleiskieliset virkkeet yleensä, jotka ovat käytetystä osakorpuksesta riippuen 11 ja 12 sanan paikkeilla (Heikkinen ym., 2001, 14). Sanojen pituudet merkeissä mitattuna sen sijaan ovat jokseenkin saman pituisia, vaikka EU-korpuksessa sanat ovatkin keskimäärin hieman pidempiä: Parolen korpuksessa keskimääräinen sanapituus merkeissä on 8,5 merkkiä, Oulun korpuksessa taas 7,5 merkkiä (emt. 14). Yli 9 merkkiä sisältäviä sanoja esiintyy EU-korpuksessa seuraavasti:



Kuva 1: Keskimääräistä pidempien sanojen määrä EU-korpuksessa

Prosentuaalisesti korpuksesta yli 9 merkkiä pitkiä sanoja on siis noin 36,9 prosenttia, mikä saattaa olla yksi syy siihen, että sanat ovat keskimääräistä pidempiä. Toisaalta se, että sanojen pituuden keskihajonnan ja keskimäärän välillä on näin suuri ero saattaa puolestaan johtua siitä, että eripituisia sanoja käytetään vaihtelevasti, ja että myös lyhyitä sanoja käytetään runsaasti.

Saatujen tulosten varjossa näyttäisikin siltä, että Euroopan unionin suomenkielisissä lakiteksteissä käytetään vaihtelevasti eripituisia sanoja, joskin pitkien, yli 9 merkkiä sisältävien sanojen prosentuaalinen määrä vaikuttaa suurelta. Myös virkkeiden pituudet ovat huomattavan pitkiä yleiskielisiin vertailukohteisiin nähden.

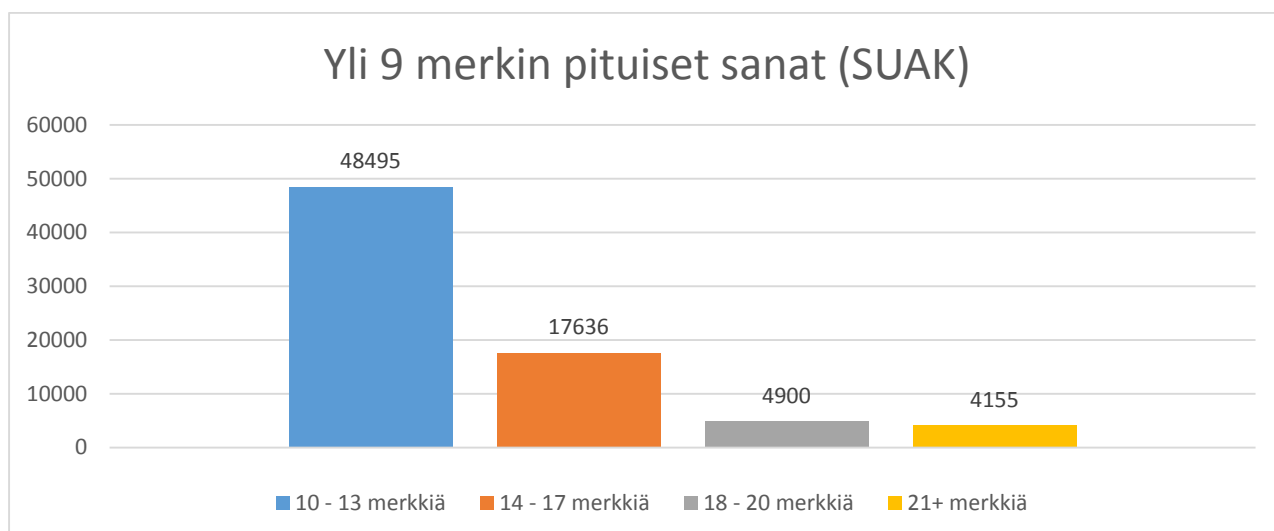
6.1.2 SU-korpus

Suomen säädöksistä koostuva, kaikkiaan 200 680 sanaa sisältävässä korpuksessa puolestaan on 22 964 erillistä sanaa, joka on yli 2000 enemmän kuin EU-korpuksessa, vaikka SU-korpus onkin sanamäärältään pienempi. Virkkeitäkin on SU-korpuksessa EU-korpusta enemmän, hieman yli 8000 kappaletta.

Taulukko 2: Sana- ja virketilastoja (SU-korpus)

sanojen keskim. pituus (merkeissä)	sanojen pituus (keskihajonta) (merkeissä)	virkkeiden määrä	virkkeen keskim. pituus (sanoissa)	virkkeen keskim. pituus (keskihajonta) (sanoissa)
8,20	4,93	8 191	22,45	24,20

Kuten taulukosta voidaan todeta, ovat sanat siis keskimäärin hieman lyhempiä kuin EU-korpuksessa, minkä lisäksi prosentuaalinen ero sanojen pituuden keskiarvon ja keskihajonnan välillä on hieman pienempi kuin EU-korpuksessa (SU-korpuksen 40 prosenttia vs. EU-korpuksen 44 prosenttia), joten vaikka tässäkin korpuksessa sanojen pituus vaihtelee, ei se vaihtele kuitenkaan yhtä paljon kuin EU-korpuksessa. Tämän lisäksi prosentuaalinen ero virkkeen pituudessa keskiarvolla ja keskihajonnalla mitattuna on huomattavasti pienempi kuin EU-korpuksessa (SU-korpuksen n. 7,9 prosenttia vs. EU-korpuksen n. 17,6 prosenttia), mistä voidaan päätellä, että SU-korpuksessa virkkeiden pituudet eivät vaihtelee yhtä paljon kuin EU-korpuksessa. Myös virkkeiden pituudet ovat kummallakin arvolla mitattuna lyhempiä: tästä voidaan siis päätellä, että Euroopan unionin teksteissä käytetään pitempiä virkeitä, vaikkakin myös SU-korpuksen virkkeet ovat huomattavasti pitempiä kuin yleiskielessä yleensä (Heikkinen ym., 2001, 14). Sanojen pituudet ovat sen sijaan hieman lyhempiä kuin Parolen korpuksessa (8,5), mutta toisaalta pitempiä kuin Oulun korpuksessa (7,5). Yli 9 merkkiä pitkiä sanoja esiintyy SU-korpuksessa seuraavasti:



Kuva 2: Keskimääräistä pidempien sanojen määrä SU-korpuksessa

Prosentuaalisesti SU-korpuksesta yli 9 merkkiä pitkiä sanoja on siis n. 37,5 prosenttia korpuksesta, joka on kenties hieman yllättäen aavistuksen enemmän kuin EU-korpuksessa. Näin ollen voidaan päätellä, että alun perin suomeksi kirjoitetuissa lakiteksteissä on yhtä paljon tai jopa hieman enemmän pitkiä sanoja ja yhdyssanoja kuin Euroopan unionin teksteissä. Toisaalta, koska sanojen keskimääräinen pituus on tästä huolimatta hieman pienempi SU-korpuksessa, viittaisi tämä siihen, että siinä käytetään myös enemmän keskimääräistä lyhempiä sanoja. Myös korpuksen tarkempi tarkastelu tukee tätä päätelmää: SU-korpuksessa on sen pienemmästä koosta huolimatta määrällisesti enemmän juuri kaksi- ja kolmikirjaimisia sanoja.

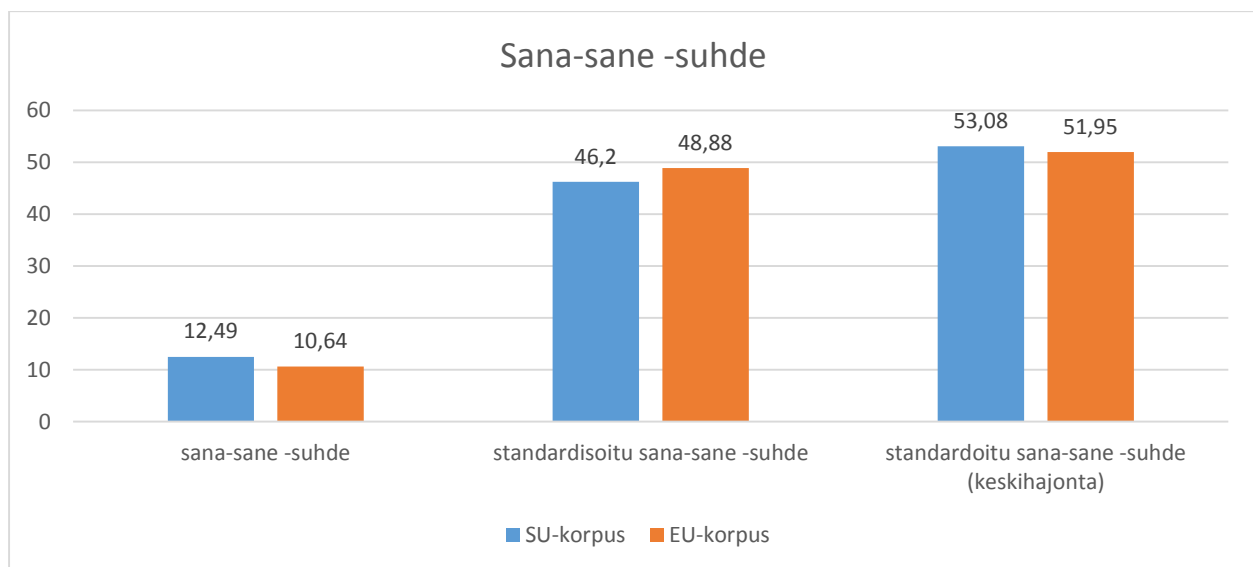
Tuloksista voidaankin päätellä, että vaikka ero sanojen pituudessa ja pitkien sanojen määrässä on molemmissa korpuksissa suurin piirtein sama, on suurin ero korpuksien välillä lauseiden määrässä ja niiden keskimääräisissä pituuksissa. Myös Sara Laviosa on vertaillut käännettyjen ja alkuperäiskielisten tekstien keskimääräisiä lausepituuksia ja tullut samaan lopputulokseen: siis siihen, että käännettyjen tekstien lauseet ovat pidempiä (Laviosa 1998, 561). Tämä saattaisikin viitata Andrew Chestermanin lähdeuniversaaliksi luokittelemaan *pidentymiseen* (*lengthening*), jonka mukaan käännöksillä on tapana olla alkuperäistekstejä pidempiä (Chesterman 2004, 40), mutta koska vertailun kohteena ovat komparatiiviset korpuksia, jolloin tulisi keskittyä kohdeuniversaalien tutkimiseen, olisi pidentymisen käännösuniversaalihypoteesista puhuminen tässä yhteydessä ongelmallista, sillä mikäli korpus ei sisällä lähdetekstejä, on lähdeuniversaalien tutkiminen korpuksessa mahdotonta (emt, 43; ks. myös luku 3.). On kuitenkin kiinnostavaa huomata, että

Svetlana Probirskaja sai samankaltaisia tuloksia, vaikkakin hän epäili niiden johtuvan Venäjän kielen vaikutuksesta ja suomalaisissa käännöksissä esiintyvien funktiosanojen määrästä (2009, 135).

On kuitenkin otettava huomioon, että Euroopan unionissa on käytössä niin sanottu ”pistesääntö”, joka vaatii, että virkkeiden määrä on jokaisessa lakiteksti eri kieliversiossa oltava sama. Tämä taas saattaa aiheuttaa ongelmia käännettäessä varsinkin, mikäli suomeksi jokin asiakokonaisuus ilmaistaisiin luontevammin kahdessa erillisessä virkkeessä. (Piehl 2003, 19–20.) Näin ollen kyseessä on mitä suuremaalla todennäköisyydellä instituution normien aiheuttama kielellinen ilmiö, eikä niinkään mikään käännösuniversaaliteorian hypoteeseista.

6.2 Sana-sane -suhteet sekä standardoidut sana-sane -suhteet

Korpuksien sana-sane -suhteita vertailtaessa saadaan mielenkiintoisia tuloksia: sana-sane -suhteissa on huomattavia eroja laskutavasta riippuen, mutta kaikki luvut ovat kuitenkin molemmissa korpuksissa hyvin samansuuntaisia (kuva 3).



Kuva 3: Sana-sane -suhteet korpuksissa

Ensimmäinen arvo, keskimääräinen sana -sane-suhde, lasketaan koko korpuksesta, ja sen matala arvo kertoisi tässä tapauksessa siis yksinkertaisesta kielestä, jossa on vähän vaihtelua. Standardoitu sana-

sane -suhde sekä standardoidun sana-sane -suhteen keskihajonta sen sijaan ovat huomattavan korkeita, mikä viittaa runsaaseen vaihteluun ja näin ollen monipuoliseen kieleen. Tulokset ovat siis jokseenkin ristiriitaisia. Selitys tähän löytyy korpuksista itsestään, mutta mahdollisesti myös instituutionaalisille teksteille tyypillisistä piirteistä.

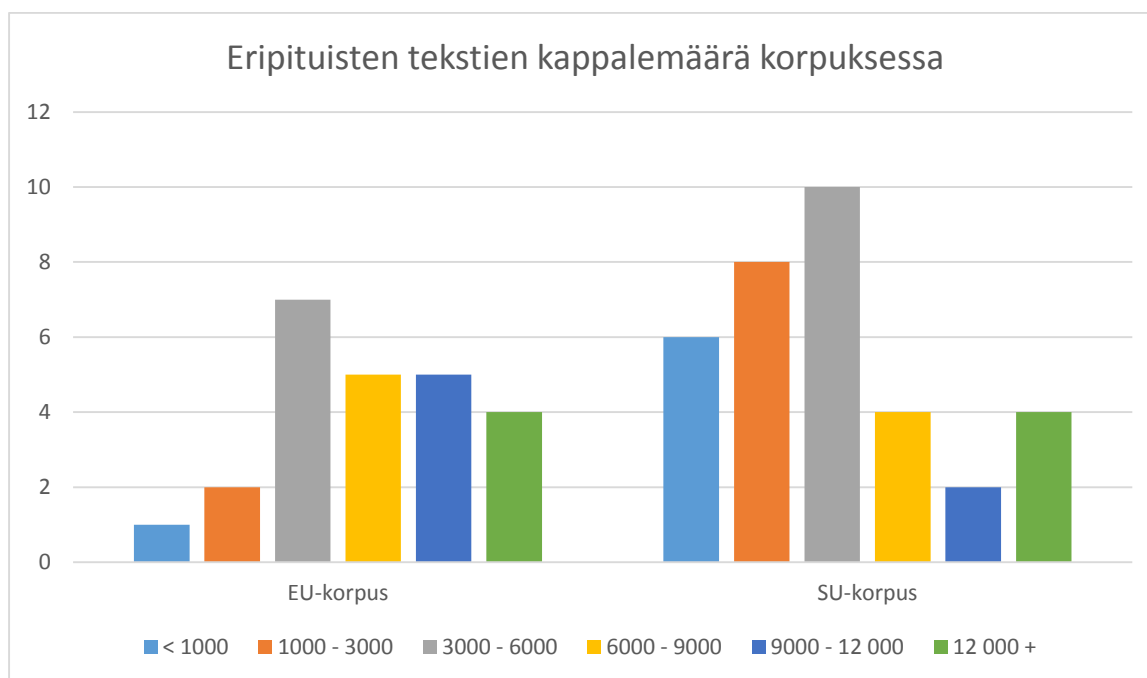
Mitä suurempi korpus, sitä useammin tietyt usein käytetyt sanat, kuten *olla*-verbin eri muodot sekä konjunktiot kuten *että*, *ja*, sekä *tai* siinä esiintyvät. Tämä puolestaan laskee sana-sane -suhdetta ja saa korpuksessa olevat tekstit vaikuttamaan yksinkertaiselta. Lisäksi korpuksien matalia sana-sane -suhteita saattaa selittää instituutionaalisen tekstilajin yleinen piirre: toisto, joka on yksi virkakielen piirteistä (Hiidenmaa 2000, 46). Jos sama lause tai lausekkeen osa toistuu eri teksteissä jatkuvasti, nostaa se korpuksien sana-sane -suhteita. Korpuksien frekvenssilistojen tarkastelu tukee osittain tätä päätelmää: kummasakin korpuksessa esiintyy eniten käytettyjen sanojen joukossa lakikielelle tyypillisiä sanoja. Esimerkiksi SU-korpuksessa yleisimmin käytettyjen sanojen listalle nousee sanoja kuten *momentissa*, *säädetään*, sekä sanan *laki* eri muodot. EU-korpuksessa puolestaan sanat *artiklan*, *jäsenvaltio* sekä sanan *asetus* eri muodot ovat hyvin yleisiä. Joskin tämä päätelmä on ongelmallista tehdä pelkästään korpuksien sanalistoja tarkastelemalla, sillä erikoisalan korpuksessa erikoisalalle tyypilliset ilmaukset ja termit ilmenevät luonnollisesti teksteissä useammin.

Sama ilmiö toistuu kuitenkin myös yksittäisten lakitekstien sanalistoissa, joissa korostuvat luonnollisesti kyseisen lain aihepiiriin kuuluvat sanat, joskin aineiston pieni määrä vaikuttaa myös tähän ilmiöön. Tämän lisäksi on luonnollista, että aihepiirille ja tekstilajille tyypilliset piirteet toistuvat erikoisalan tekstejä sisältävässä korpuksessa, kuten yllä mainittiin. On kuitenkin myös mielenkiintoista, että esiin nousee juuri aihepiiriin liittyviä substantiiveja tekstistä riippuen, kuten esimerkiksi *passi*, *lääke* ja *tuomioistuin*. Tätä saattaisi selittää Hiidenmaan huomio virkatekstien luonteesta: hänen analyysinsä mukaan lain ja hallinnon teksteissä ydinsanan yhteydessä esiintyy runsaasti määritteitä. Erityisesti substantiivimääritteitä on runsaasti, ja juuri nämä tekevät ydinsanaan viittaamisen pronominilla vaikeaksi. (Hiidenmaa 2000, 46.) Siis, koska koheesiota ei voida rakentaa pronomien varaan, rakentuu koheesio toiston varaan, mikä käytännössä johtaa siihen, että jotain lauseketta, fraasia, ydinsanaa tai ydinsanan määritettä toistetaan tekstissä (emt., 46–47). Toisaalta myös demonstratiivi- sekä relatiivipronomit usein käytettyjen sanojen joukkoon, vaikka niiden käyttö yksittäisissä teksteissä on sanalistojen mukaan vähäisempää ja vaihtelee tekstistä toiseen. Tämä voisi siis viitata siihen, että toistoa tapahtuu myös yksittäisten tekstien tasolla, joskaan ilman laajaa kvalitatiivista analyysia on vaikeaa tehdä ratkaisevia johtopäätöksiä.

Toisaalta niiden sanojen prosentuaalinen määrä, joita käytetään vain kerran, on EU-korpuksessa 45,79 prosenttia, ja SU-korpuksessa puolestaan 46,30 prosenttia. Tämä saattaa johtua korpuksen pienestä sanamäärästä ja kokoamistavasta. Kuten luvussa 5 on esitetty, on kumpaankin korpuksen kerätty tekstejä mahdollisimman useasta eri lain alueesta markkinataloutta ja ympäristöä koskevista laeista rikoslakiin. Lain aihepiiri taas puolestaan vaikuttaa sanastoon, jota siinä käytetään: korpus kokonaisuudessaan sisältää siis hyvin paljon eri sanoja, mutta niitä saatetaan käyttää vain muutamissa teksteissä. Tämän pitäisi periaatteessa nostaa korpuksien sana-sane -suhteita, mutta kun otetaan huomioon, että osa sanoista toistuu korpuksessa tekstistä toiseen useampaan kertaan, jää näiden yksittäisten sanojen vaikutus korpuksen pieneksi. Kun kaikki tämä otetaan huomioon tutkittaessa standardoituja sana-sane -suhteita ja niiden keskihajontaa on loogista, että nämä arvot ovat korkeampia, sillä ne lasketaan aina 1000 sanan välein (ks. luku 4.3), joten edellä kuvatun kaltainen toisto vaikuttaa niihin luonnollisesti vähemmän.

Kun sen sijaan vertaillaan korpuksista saatuja arvoja keskenään, huomataan, ettei niiden välillä ole juuri vaihtelua (kuva 3): vaikka keskimääraisten sana-sane -suhteiden välillä on 14,8 prosentin ero, on standardoidun sana-sane -suhteen ero vain 5,5 prosenttia, ja niiden keskihajonnan ero on vain hieman yli 2 prosenttia. Näin ollen näyttäisi siltä, että suhdelukujen välinen vaihtelu selittyisi normaalilla vaihtelulla, eikä niiden välillä näytä olevan mitään huomattavia eroja. Tämä puolestaan tarkoittaisi sitä, etteivät saadut tulokset tue käänösuniversaalihypoteesia käännösten kielen yksinkertaistumisesta.

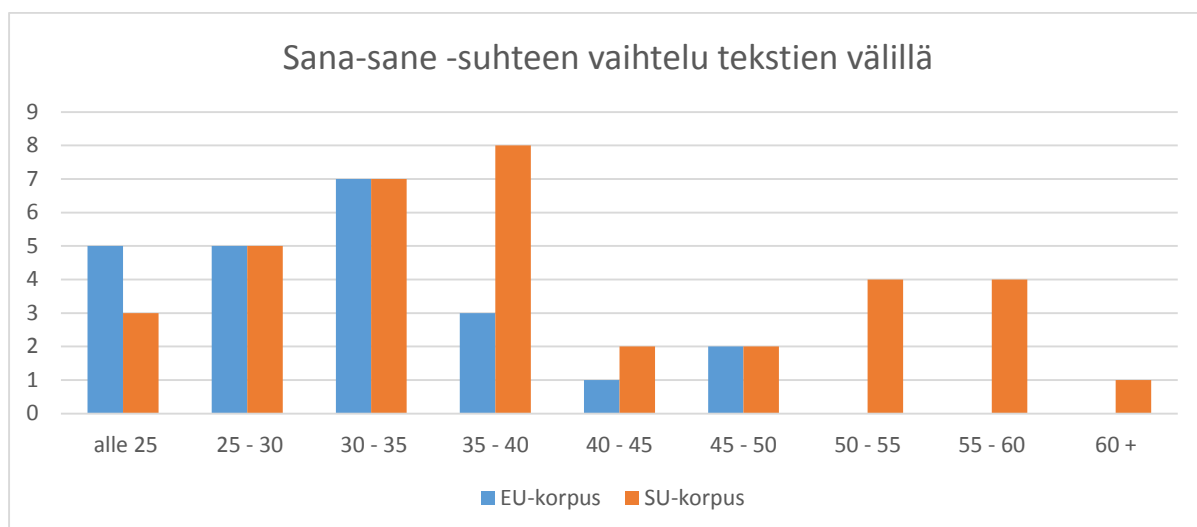
Toisaalta, mikäli yksittäisillä teksteillä olisi huomattava vaikutus korpuksessa, saattaisi tämä vääristää tulosta, minkä vuoksi onkin hyvä tutkia, minkä pituisia tekstejä korpuksessa on, ja kuinka paljon. Kuten kuvasta 4 käy ilmi, näissä onkin vaihtelua niin korpuksien sisällä kuin niiden keskenkin.



Kuva 4: Sanamäärältään eripituisten tekstien määrä korpuksissa

Kuvasta käy siis ilmi, että vaikka tekstien määrä eroaa selvästi, painottuvat niissä olevien tekstien pituudet eri tavalla korpuksissa: Suomen säädösteksteistä koostuvassa SU-korpuksessa on huomattavasti enemmän hieman lyhempiä, alle 3000 sanaa pitkiä tekstejä kuin EU-korpuksessa. EU-korpuksessa taas on enemmän pitempiä, yli 6000 sanan tekstejä kuin SU-korpuksessa. Hyvin pitkiä, yli 12 000 sanan tekstejä on sen sijaan molemmissa korpuksissa yhtä paljon, joskin EU-korpuksen pisin teksti, joka on 29 456 sanaa pitkä, on yli 10 000 sanaa lyhempi kuin SU-korpuksen pisin teksti, joka on 39 485 sanaa pitkä. Näiden tekstien poistaminen korpuksesta ei kuitenkaan juuri vaikuta korpuksesta saatuihin tuloksiin: kummassakaan korpuksessa tulokset eivät juuri muutu, joskin molempien korpuksien sana-sane -suhteet nousevat noin yhden yksikön verran. Tästä huolimatta molempien korpuksien standardoitu sana-sane -suhde sekä sen keskihajonta pysyvät kuitenkin lähes täsmälleen samoissa lukemissa. Näin ollen kyseisten tekstien ei pitäisi vääristää korpuksesta saatuja tuloksia.

Tutkimalla yksittäisten tekstien sana-sane -suhteiden jakautumista korpuksista selviää, että sana-sane -suhteet vaihtelevat huomattavasti tekstien välillä. Samalla selviää kuitenkin myös, että SU-korpuksissa näyttäisi siltäkin olevan enemmän tekstejä, joilla on korkeampi sana-sane -suhde, ja jotka näin ollen olisivat kieleltään monipuolisempia (kuva 5), mikä tukisi yksinkertaistumisen käännösuniversaalihypoteesia.



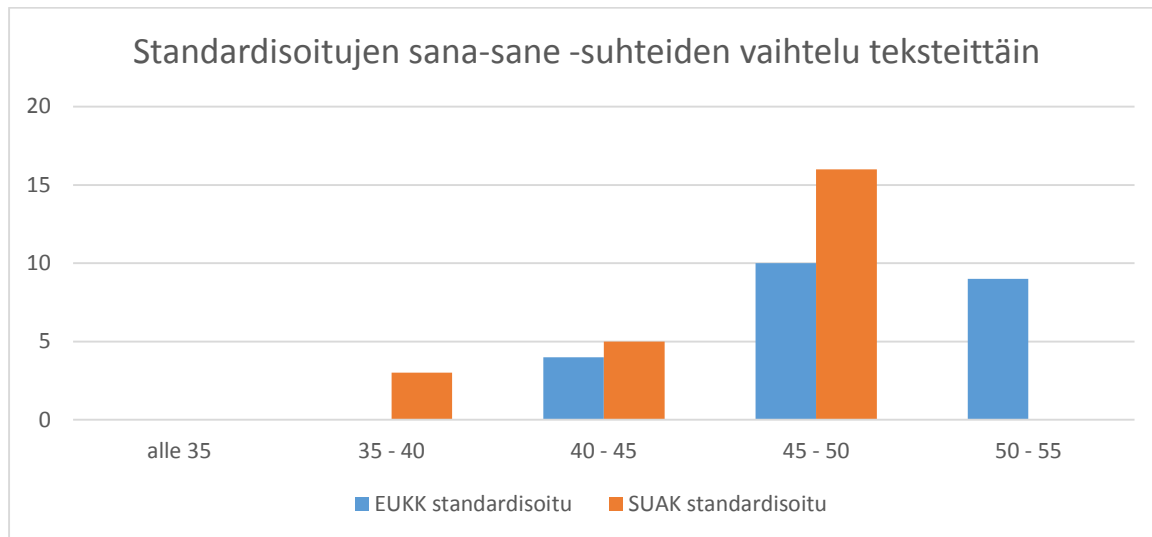
Kuva 5: Sana-sane -suhteet korpuksien teksteissä ja näiden tekstien lukumäärät korpuksissa

Kuvan mukaan Euroopan unionin käännöskorpuksessa ei ole yhtään tekstiä, jonka sana-sane -suhde ylittäisi 50 arvon. Tämä on kuitenkin hieman harhaanjohtava huomio. Tuloksia korpuksessa tarkasteltaessa paljastuu nopeasti, että Suomen säädösteksteistä koostuvassa korpuksessa korkeita arvoja saavat lyhyet, noin 1000 sanaa sisältävät tekstit, joita on SU-korpuksessa huomattavasti enemmän (kuva 4) ja näin ollen tekstien pituus vaikuttaa näihin tuloksiin. Tämä saattaa selittää osittain myös sen, miksi SU-korpuksen sana-sane -suhde on lähes kaksi yksikköä EU-korpuksen sana-sane-suhdetta suurempi (kuva 3).

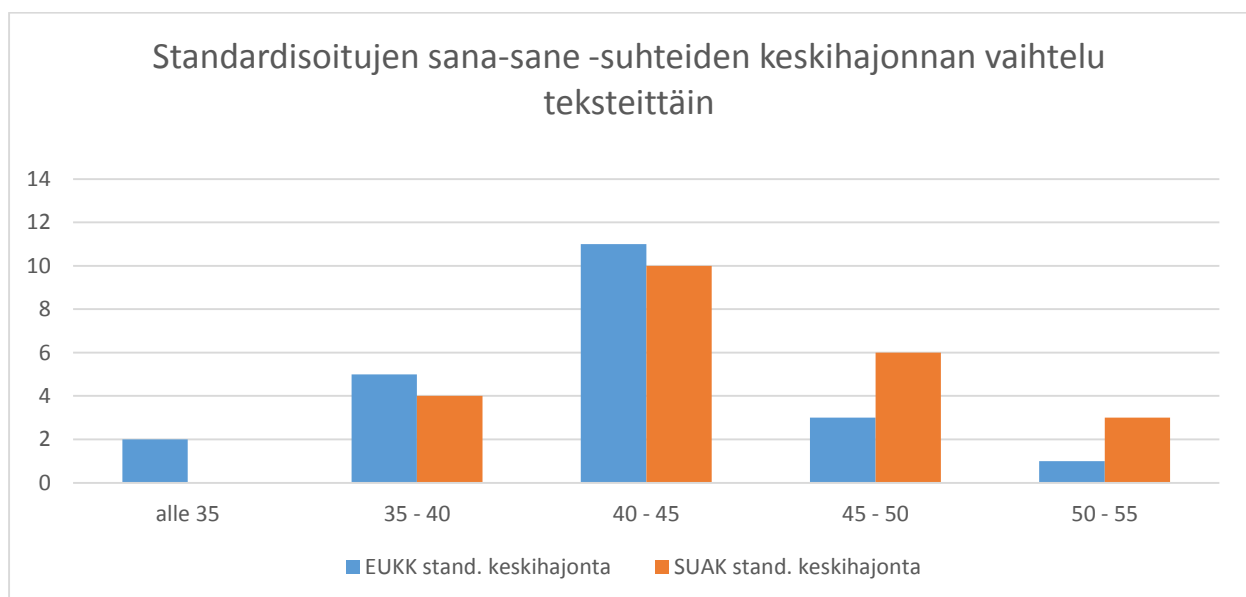
Näistä eroista huolimatta korpuksien keskinäiset standardoidut sana-sane -suhteet sekä niiden keskihajonta ovat kuitenkin hyvin samanlaisia. Tämä johtuu osittain siitä, että koska WordSmith Tools laskee standardoidun sana-sane-suhteen sekä sen keskihajonnan aina 1000 sanan välein, joten kaikkein lyhimpiä tekstejä ei edes oteta huomioon näissä laskuissa. Tästä huolimatta näiden tekstien poisjättäminen korpuksista ei vaikuta suuresti saatuihin tuloksiin. Korpuksen sana-sane -suhde pysyy lähes samana: se nousee vain yksikön kymmenesosalla. Näin ollen voidaan olettaa, ettei kyseisillä teksteillä ole suurta vaikutusta korpuksessa.

Kun taas tutkitaan standardoitujen sana-sane -suhteiden vaihtelua teksteittäin, saadaan edellisen kaltaisia tuloksia, joskin tällä metodilla mitattuna saadut tulokset ovat edellisen tulosten kanssa ristiriidassa. Kuten kuvasta 6 näkyy, on SU-korpuksen kieli standardoidun sana-sane -suhteella mittaamalla yksipuolisempaa, sillä siinä ei ole ainuttakaan tekstiä, jonka standardoitu sana-sane

-suhde olisi yli 50 yksikköä, ja siinä on myös enemmän tekstejä, joiden sana-sane -suhde on alle 45. Näin ollen tulos on päinvastainen kuin standardoiduilla sana-sane -suhteilla mitattuna.



Kuva 6: Standardoitujen sana-sane -suhteiden vaihtelu teksteittäin eri korpuksissa



Kuva 7: Standardoitujen sana-sane -suhteiden keskihajonnan vaihtelu teksteittäin eri korpuksissa

Toisaalta, kun taas vertaillaan standardisoitujen sana-sane-suhteiden keskihajontaa (kuva 7), on tilanne päinvastainen, ja EU-korpuksen saamat arvot ovat taas pienempiä. WordSmith Tools ei ole kuitenkaan pystynyt laskemaan standardisoitujen sana-sane -suhteiden keskihajontaa kolmesta SU-korpuksen tekstistä, ja tähän puolestaan vaikuttaa se, että se lasketaan uudelleen aina 1000 sanan

jälkeen. Näin ollen arvoa ei ole saatu teksteistä, jotka ovat vain hieman yli 1000 sanan pituisia, mikä saattaa vaikuttaa tulokseen. Näin ollen standardoidun sana-sane -suhteen laskemiseen käytetyn jakson pituus saattaisi vaikuttaa tuloksiin, mutta sen laskeminen 800:aan ei aiheuta suuria muutoksia (taulukko 3): vaikka arvot vaihtelevat, ovat ne vieläkin samansuuntaisia molemmissa korpuksissa, eikä niissä ole radikaalia eroa, ja niiden väliset prosentuaaliset erot jäävät alle kolmeen prosenttiin. Lisäksi huomataan taas, että kun standardoitu sana-sane -suhde on EU-korpuksessa korkeampi kuin SU-korpuksessa, on tilanne päinvastoin keskihajontaa tarkastellessa.

Taulukko 3: Standardoidut sana-sane -suhteet 800 sanan jaksoissa laskettuina

	sana-sane -suhde (stand.)	sana-sane -suhde (stand. keskih.)
EU-korpus	51,36	49,69
SU-korpus	48,41	51,25

Jos SU-korpukselta taas poistetaan kaikki alle 1000 sanan tekstit ja standardoitu sana-sane -suhde lasketaan 800 sanan välein, ovat tulokset muuten lähes samat, mutta standardoitu sana-sane -suhde on hieman alhaisempi 48,39. Tuloksissa ei siis tapahdu mitään radikaaleja muutoksia, ja ne ovat kaiken kaikkiaan hyvin lähellä toisiaan.

Tämän perusteella onkin todennäköistä, että vaikka vaihtelua on havaittavissa, on se suurimmaksi osaksi samansuuntaista – eli tekstikohtainen vaihtelu korpusten sisällä on kummassakin korpuksessa suurin piirtein samanlaista, eikä yhdellä tekstillä pitäisi olla huomattavaa vaikutusta sana-sane -suhdetta koskeviin tuloksiin. Tämä käy hyvin ilmi esimerkiksi kuvasta 7: vaikka eroja on havaittavissa, eivät ne kuitenkaan ole määrällisesti erityisen suuria, joskin tekstien vähäinen määrä tekee todellisesta vertailusta vaikeaa. Kaiken kaikkiaan saatujen tulosten perusteella vaikuttaisi kuitenkin siltä, että korpuksella todella ovat sanastonsa monipuolisuuden suhteen niin samankaltaisia, ettei saatujen tulosten voida sanoa tukevan käännösuniversaaliteorian yksinkertaistumishypoteesia.

6.3 Sanaluokkajakauma ja leksikaalinen tiheys

Molempiin korpuksiin tehtiin sanaluokkajäsennys käyttämällä TreeTagger-ohjelmaa, jonka tarkkuudeksi saatiin manuaalisen laskennan tuloksena n. 89,5 prosenttia¹. Kuten kuvasta 9 käy ilmi, ei korpus ole kyennyt luokittelemaan 7–8 prosenttia kaikista sanoista (kategoria NON-TWOL). Tästä huolimatta korpuksen tulokset näyttävät olevan ainakin jollain tavalla vertailtavissa. Tulosten pohjalta näyttäisikin siltä, että ainoa merkittävä ero korpusten kesken on adjektiivien määrässä, joita on jokseenkin enemmän EU-korpuksessa. Mikään muut taulukon lukemat eivät myöskään anna aihetta epäillä, että esimerkiksi kuvailevien partisiippien tai adjektiiviattribuuttien suosiminen adjektiivien sijaan selittäisi tätä ilmiötä.

Taulukko 4: Korpusten sanaluokkajakaumat

EUKK		SUAK	
Sanaluokka	% määrä	Sanaluokka	% määrä
Subst.	45,1	Subst.	44,9
Verb.	11,0	Verb.	10,8
Partis. ²	7,1	Partis.	6,9
Adj.	8,2	Adj.	4,9
Konj.	7,0	Konj.	8,2
Num.	5,0	Num.	7,3
Pron.	4,6	Pron.	3,4
Adv.	2,3	Adv.	2,3
NON-TWOL	6,8	NON-TWOL	8,2
muut	3,0	muut	3,0

¹ Satunnainen 200 luokitellun sanan otos, virhemarginaali 95 % varmuudella $\pm 7,7$ prosenttiyksikköä.

² Partisiipit on merkitty kaaviossa omaksi luokakseen, sillä ne voidaan kontekstista riippuen tulkita joko verbeinä tai adjektiiveina.

Saatujen tulosten perusteella EU-korpuksen leksikaaliseksi tiheydeksi saadaan 73,7 prosenttia, ja SU-korpuksen leksikaaliseksi tiheydeksi 69,95 prosenttia. Tulos on jokseenkin samansuuntainen Pirjo Hiidenmaan saamien tulosten kanssa, joskin Hiidenmaan aineiston leksikaalinen tiheys oli 74–77, eli hieman korkeampi (2000, 44). Sanaluokkajakauman suhteen on myös nähtävissä eroja sekä Parole-korpuksen että Hiidenmaan saamien tuloksien kanssa. Substantiiveja on vertailtavissa korpuksissa noin viisi prosenttiyksikköä vähemmän kuin Hiidenmaan aineistossa, mutta myös ero verbien prosentuaalisissa määrissä on huomattava: Hiidenmaan aineistossa verbejä oli 16 prosenttia kaikista sanoista (2000, 44) kun sekä EU-korpuksessa että SU-korpuksessa vastaava summa oli noin 11 prosenttia. Parole-korpuksen verrattuna erot verbien määrässä ovat vielä huomattavampia: jopa Parolen sanomalehtiteksteistä koostuva osakorpuksesta, jossa verbien määrä on prosentuaalisesti pienin, yli 20 prosenttia sanoista on verbejä (Heikkinen ym., 2001, 14).

Koska määrät ovat suurin piirtein samat sekä EU- että SU-korpuksissa ja Hiidenmaan oma tutkimus keskittyi esityslistatekstien analyysiin, jolla on hieman erilaiset kielelliset konventiot, voidaan prosentuaalisesti matalampi verbimäärä tulkita myös lakikielelle ominaiseksi leksikaaliseksi piirteeksi. Toisaalta tämä päätelmä saattaisi olla hieman outo siinä valossa, että laki on sekä Hiidenmaan että Caon mukaan performatiivista, jonka luonteeseen loogisesti kuuluisi verbien ahkera käyttö. Korpuksissa esiintyvät pitkät virkkeet ja Hiidenmaan mainitsema toisto sekä substantiivien rinnastus (2000, 51) saattavat tietenkin laskea verbien prosentuaalista määrää.

Toisaalta voidaan myös kysyä, kuinka vertailukelpoisia tulokset näiden aikaisempien korpuksien tulokset ovat näiden tuloksien kanssa, sillä kummakaan korpuksella ei sisällä tekstejä, jotka kuuluisivat täsmälleen samaan tekstilajiin tässä tutkimuksessa käytettyjen korpusten tekstien kanssa. Hiidenmaan aineisto koostuu esityslistateksteistä, joiden ei voida välttämättä katsoa kuuluvan samaan tekstilajiluokkaan kuin lakitekstien, vaikka molemmat kuuluvatkin laajempaan, lain ja hallinnon tekstien kokonaisuuteen. Parole-korpuksen käyttö vertailukohtana on kenties vielä ongelmallisempaa, sillä siinä olevat tekstit eivät edes kuulu lain ja hallinnon tekstilajin piiriin.

Sen sijaan se, että käännöskielisen korpuksen leksikaalinen tiheys on kohdekielistä korpusta korkeampi, on ristiriidassa Sara Laviosan saamien tulosten kanssa (1998, 561). Tämä saattaa olla korpuksen sekä lauseenosajäsennyksen aiheuttama ongelma: SU-korpuksessa on jäänyt prosentuaalisesti enemmän sanoja kategorisoimatta, mikä saattaa aiheuttaa eron tuloksissa. Tämän

lisäksi lauseenjäsennyksen luotettavuus on otettava tuloksia tulkittaessa huomioon (ks. alaviite 1). Toisaalta on myös mahdollista, että korpus on kategorisoimatta jääneistä sanoista huolimatta oikeassa ja että yksinkertaistumishypoteesi ei tässäkään yhteydessä pidä paikkaansa.

Näin ollen voidaan sanoa, ettei saatujen tulosten perusteella ole saatu tuloksia, jotka tukisivat käännösuniversaaliteorian yksinkertaistumishypoteesia sanastollisella tai leksikaalisella tasolla.

7. PÄÄTELMÄT

Edellä mainittujen tulosten perusteella voidaan todeta, etteivät tämän tutkimuksen tulokset tue käännösuniversaaliteorian yksinkertaistumishypoteesia. Näin ollen saadut tulokset ovat ristiriidassa aikaisemman tutkimuksen kanssa (esim. Laviosa 1998, Nevalainen 2005), mutta koska kyseiset tutkimukset on tehty erilaisella aineistolla, voi kyseessä olla hyvin mielenkiintoinen löytö. Tutkimuksen pohjalta tehtiin myös muita mielenkiintoisia havaintoja lakikielen käännöksistä.

Vaikka tutkimuksen perusteella korpuksien sana-sane -suhteissa on eroja, eivät ne tarpeeksi huomattavia, jotta niiden voitaisiin tulkita tukevan yksinkertaistumishypoteesia. Kun otetaan kuitenkin huomioon aikaisempi tutkimus ja sen päinvastaiset tulokset, on syytä kysyä, mistä eriävä tulos on peräisin. Eräs mahdollinen selitys tälle ilmiölle on tutkimusmateriaali: kuten luvussa 2 todettiin, ovat lain ja hallinnon tekstit luonteeltaan normatiivisia, ja tästä syystä voidaan olettaa, että myös niiden kääntäminen on hyvin normitettua, varsinkin kun on kyse todistusvoimaisista lainkäännöksistä. Olisiko siis mahdollista, että käännösuniversaaleiden esiintymiseen vaikuttavat myös käännösnormien puute? Lain ja hallinnon tekstien kääntämistä määrittävät huomattavasti tiukemmat normit kuin esimerkiksi kaunokirjallista kääntämistä, joten olisiko mahdollista, että nämä tiukat normit vaikuttavat siihen, miten, kuinka paljon ja millaisia käännösuniversaaleja käännöksissä esiintyy. Saattaakin olla mahdollista, että lain ja hallinnon teksteillä on omat käännösuniversaalinsa, ja yksinkertaistuminen vain ei ole yksi niistä.

Saadut tulokset ovat myös kiinnostavia ekvivalenssin käsitteen näkökulmasta. Luvussa 2.2 mainittu ekvivalenssin korostuminen lakikäännöksissä saattaa tarkoittaa saatujen tulosten näkökulmasta myös sitä, että ekvivalenssia ei ole pelkästään lähde- ja kohdetekstien välillä, vaan myös kohdetekstin ja kohdekulttuurin välillä. Toisin sanoen, onko mahdollista, että koska käännettävät tekstit tulevat osaksi kohdekulttuurin lain ja hallinnon tekstien tekstilajiketjua, pyritään näiden tekstien piirteet ja niissä esiintyvät elementit ottamaan huomioon käännösprosessin aikana esimerkiksi käyttämällä Suomen säädöksiä rinnakkaisteksteinä käännösprosessin aikana? Piehl on maininnut, että vaikealukuisista EU-säädöksistä koituu hankaluuksia jäsenmaiden hallinnolle (2003, 19). Onko siis mahdollista, että tätä on pyritty vähentämään juuri pyrkimällä mahdollisimman samantapaiseen ilmaisuun suomalaisten lakitekstien kanssa? Voitaaisiinko siis Erich Prunčin termiä käyttäen puhua *käännöskulttuurista* (*Translationskultur*, käännös Pekka Kujamäki 2013, 33), joka pyrkii

käännösratkaisuihillaan ja -normeillaan ekvivalenssiin Euroopan unionin lakitekstien suomennosten ja suomalaisen lain ja hallinnon tekstikulttuurin välillä?

Tämä taas olisi ristiriidassa Kaisa Koskisen näkemyksen kanssa, jonka mukaan tekstiversioiden välinen ekvivalenssi olisi kohdetekstin ja kohdekulttuurin välistä ekvivalenssia tärkeämpää (2011, 58; ks. myös luku 2.2), mutta tämä ei näytä realisoituvan omassa tutkimuksessani ainakaan tässä tutkimuksessa tarkastelluilla kielen tasoilla. Euroopan unionin teksteille tyypilliset fraasit, ilmaisut ja termit, jotka eivät ole kohdekulttuurille tyypillisiä, ovat tietenkin oma kysymyksensä. Myös se, miten Koskisen mainitsemat tekstinvälisen koheesion keinot instituutionaalisessa kääntämisessä, kuten tieto- sekä termipankit ja käännösmuistit, ovat mahdollisesti vaikuttaneet tulokseen, on hyvin mielenkiintoinen kysymys. Onko esimerkiksi mahdollista, että käännösmuisteihin on käännetty tekstejä käyttäen suomenkieliselle lakitekstille ja Suomen lakikielen kulttuurille tyypillisiä rakenteita ja muotoiluja? Tämä saattaisi selittää tulosten yhtenäisyyttä.

Sana-sane -suhteen matala arvo koko korpuksesta laskettuni oli sen sijaan yhdenmukainen Hiidenmaan löytöjen kanssa, jonka mukaan toisto on yksi viranomaistekstien tunnusomaisista piirteistä (2000, 46). Tämä piirre mainitaan myös Oikeusministeriön Lainkirjoittajan oppaassa, jonka mukaan vakioilmaisuiden käyttäminen kuuluu lakikielen luonteeseen (2013, 439), ja mikäli sama vakioilmaus tai sana esiintyy runsaasti useammassa lakitekstissä, laskee tämä koko korpuksesta laskettua sana-sane -suhdetta. Tämän lisäksi sana-sane -suhteen matalampaan arvoon standardoituihin arvoihin nähden vaikuttaa luonnollisesti korpuksen koko.

Toinen analyysin pohjalta tehty mielenkiintoinen löytö on käännöskorpuksen pidemmät virkkeet, jotka ovat samassa linjassa aikaisempien löytöjen kanssa (Laviosa 1998, Probirskaja 2009). Laviosa ja Probirskaja työskentelivät eri lähdekielistä käännettyjen tekstien parissa: suurin osa Laviosan käännöskorpuksen aineistosta oli käännetty romaanisista kielistä, joista suurimman osan lähtökielenä oli ranska, kun taas Probirskajan korpuksen lähtökieli oli venäjä. Tämä viittaisi siihen, ettei ilmiö johdu välttämättä ainoastaan lähtökielen vaikutuksesta, kuten Probirskaja epäili. Toisaalta, kuten luvun 3.3 yhteydessä mainittiin, on ongelmallista väittää yksikielisestä korpuksesta tehtyjä löytöjä lähdeuniversaaleiksi, sillä lähdekieli ei ole vertailtavana.

On myös otettava huomioon jo luvun 6 analyysin yhteydessä mainittu seikka: Euroopan unionin sisäiset käännösnormit saattavat myös vaikuttaa siihen, että lauseet ovat keskimääräistä pidemmät. Tämä ei tietenkään selittäisi sitä, miksi kaunokirjallisia tekstejä tutkinut Laviosa ja valtiosopimuksia tutkinut Probirskaja ovat saaneet samanlaisia tuloksia. Löytö on siis mielenkiintoinen ja näiden tulosten perusteella onkin mielestäni aiheellista nostaa esille kysymys siitä, voidaanko pidentyminen luokitella vain lähdeuniversaaliksi. On mielestäni loogista, että tietyt käännösstrategiat, kuten eksplikointi, tekisivät lauseista pidempiä tapauksissa, kun kohdekielessä ei ole vastaavaa termiä tai ilmaisua. Se, missä määrin eksplikointi on kulttuurisidonnainen käännösstrategia, on tietenkin oma kysymyksensä.

Kolmantena mielenkiintoisena kysymyksenä esille nousee Euroopan unionin tekstejä sisältävän korpuksen adjektiivien määrä Suomen lakitekstien sisältäviin adjektiiveihin verrattuna. Havainnosta tekee erityisen mielenkiintoisen se, että korpuksissa on suurin piirtein samat määrät partisiippeja sekä substantiiveja, joten erilaiset kielelliset ilmaisustrategiat, kuten kuvailevien partisiippien tai adjektiiviattribuuttien käyttö eivät vaikuta mahdollisilta selityksiltä ilmiölle. On tietenkin mahdollista, että tämä on lähdekielen vaikutusta.

Myös EU-korpuksen hieman korkeampi leksikaalinen tiheys herättää kysymyksiä. Vaikka kyseessä onkin vain kolmen prosenttiyksikön ero, joka mahtuu virhemarginaaliin, on se yllättävä tulos kun otetaan huomioon, että aikaisempi tutkimus on viitannut nimenomaan leksikaaliseen yksinkertaistumiseen. Syitä korkeampaan leksikaaliseen tiheyteen on tämän tutkimuksen rajoissa mahdotonta spekuloida, mutta kenties tässä olisi selitys Koskisen mainitsemaalle ”outoudelle” (Koskinen 2011, 58; ks. myös luku 2.2), joka vieraannuttaa instituutionaalisten käännösten lukijoita. Toisaalta lukijoita vieraannuttava tekijä saattaa olla myös se, että virkkeet ovat keskimääräistä pitempiä. On myös mahdollista, että nämä kaksi piirrettä yhdessä vaikuttavat lukijaan vieraannuttavasti. Olisi loogista, että leksikaalisesti tiheän tekstin lukeminen, joka sisältää pitkiä virkkeitä, saattaa aiheuttaa lukijalle lukuvaikeuksia ja näin ollen myös vieraantumisen tunnetta.

Selityksiä on kuitenkin mahdollista vain spekuloida, sillä korpus pystyy vastaamaan vain kysymyksiin *mitä* ja *kuinka paljon*, ja kysymys *miksi* jääkin tässä tapauksessa valitettavasti vastaamatta. Kaiken kaikkiaan tutkimuksen pohjalta on mahdotonta tehdä ratkaisevia päätelmiä käännösuniversaalien pätevyydestä instituutionaalisten tekstien yhteydessä, joskin ainakaan

yksinkertaistumisen käännöshypoteesia tutkimuksen tulokset eivät vaikuta tukevan. Tulokset ovat itse asiassa niin yhdenmukaisia, että on esitettävä kysymys: onko mahdollista, että institutionaalisilla teksteillä on enemmän tekstilajiin kuin kieleen pohjautuvia rakenteellisia yhtäläisyyksiä jotka määrittävät niiden kääntämistä siinä määrin, etteivät perinteiset käännösuniversaalit päde niiden yhteydessä? Vai olisiko mahdollista, että niillä on omat käännösuniversaalinsa, jotka pätevät vain juuri tämän tekstilajin yhteydessä?

Tämän lisäksi voidaan nostaa myös esille Touryn käännösnormit sekä Tymoczkon idea kääntämisestä kulttuurisidonnaisena toimintana: myös instituutio voidaan nähdä omana kulttuurinaan, tai ainakin omana *työkulttuurinaan*, jossa pätevät tietyt instituution määrittämät säännöt sekä normit. Euroopan unionin erilaiset kirjoittamista ja kääntämistä koskevat oppaat ovat tästä erinomainen esimerkki: on selkeää, että Euroopan unionilla on tietty normisto, jonka mukaan sen tekstejä kirjoitetaan sekä käännetään. Suomessa samanlaisia kielenhuollon palveluja viranomaisille tarjoaa Kotimaisten kielten tutkimuskeskus, ja voidaankin kysyä, onko juurikin vahva normitus puolin ja toisin samankaltaisten tulosten syynä.

Kaiken kaikkiaan tutkimuksen pohjalta voidaan todeta, että Andrew Chestermanin kehoitus universaalien etsimisestä pienemmistä aineistoista saattaisi olla paikallaan (2010, 46). Tällä tavalla saattaisikin olla mahdollista saada uutta, valaisevaa tietoa mahdollisista käännösuniversaaleista sekä siitä, miten ne mahdollisesti käyttäytyvät eri tekstilajikonventioissa. On kenties jopa mahdollista, että tällainen pienempiin, pidemmälle erikoistuneisiin korpuksiin kohdistuva tutkimus saattaisi paljastaa joitain tekstilajikohtaisia käännösuniversaaleja. Toisaalta tarkempi tutkimus eri tekstilajikonventioiden tyypillisiin piirteisiin saattaisi myöskin asettaa käännösuniversaaliteorian tai osan sen hypoteeseista kyseenalaiseksi, tai ainakin asettaa kyseenalaistaa sen, että teoriaa kutsutaan käännösuniversaaliteoriaksi.

8. LOPUKSI

Tutkimukseni tulokset ovat antaneet aihetta lähteä tutkimaan käänösuniversaalihypoteeseja erikoisalojen sekä erikoiskielien näkökulmasta, sillä näyttäisi siltä, että ne eivät päde, tai ainakaan ole täysin samoja, eri tekstilajien välillä. Tähän saattaa olla useita syitä aina tekstilajien sisäisistä normeista tietyille tekstilajeille täysin uniikkeihin käänösuniversaaleihin asti. Tutkimukseni on myös osoittanut, että erityisesti lakitekstien korpustutkimukselle olisi kenties laajempaakin tarvetta, ja siitä löytyviä erikoispiirteitä olisikin kiinnostavaa tutkia tulevaisuudessa lisää. Saamieni tutkimustulosten perusteella näyttäisi siltä, että joko lakitekstit ovat tekstilajissaan poikkeuksellisia siltä osin, että niistä ei ole löydettävissä yhtä selkeitä käänösuniversaaleja, tai niillä on omat universaalinsa, joihin nykyiset hypoteesit eivät päde. Olisikin mielenkiintoista tutkia esimerkiksi juuri lauseiden pitenemistä myös kohdeuniversaalien osalta ja määrittää, saataisiko esimerkiksi jokin käänösstrategia tai -normi olla ilmiön taustalla.

Tulosten tulkinnassa on kuitenkin otettava huomioon käytetyn tutkimusmetodin sekä aineiston valinnan aiheuttamat rajoitukset. Pienestä korpuksesta on vaikeaa saada laajalti yleistettävää tietoa. Myös lakitekstien erikoiset piirteet, kuten muihin lakeihin ja teksteihin viittaaminen numero-, kirjain- ja välimerkkiyhdistelmiä käyttämällä on saattanut vääristää tuloksia, ja mahdollisessa tulevassa tutkimuksessa tulisikin kiinnittää huomiota siihen, miten korpustyökalut sekä muut tutkimuksessa käytetyt apuvälineet reagoivat juuri näihin tekstin erityispiirteisiin. Myös lauseenjäsennystyökalun kohdalla on otettava huomioon, ettei manuaalisesti satunnaisesta otoksesta tehty tarkkuuslaskenta ole välttämättä tarkka, ja että lisäksi sen virhemarginaali on jokseenkin suuri ± 7 prosenttia, ja että tämän lisäksi luokittelematta jääneiden yksiköiden määrä on 7–8 prosenttia. Valitettavasti tietty epätarkkuus kuuluu kvantitatiivisiin ja statistisiin menetelmiin, joskin suurempaa korpusta käyttämällä ja lauseenjäsenet manuaalisesti editoimalla saataisiin mahdollisesti tarkempia tuloksia.

Näistä puutteista huolimatta tutkimukseni on nostanut esille myös mielenkiintoisia kysymyksiä ja mahdollisuuksia uuteen tutkimukseen. Jo edellä mainittu lain ja hallinnon teksteihin kuuluvien erikoisaloja käsittelevien tekstien tutkiminen saattaisi olla kiinnostavaa ja paljastaa uutta tietoa esimerkiksi siitä, miten lakikieli omaksuu ja muokkaa erikoisalan sanastoa omiin konventioihinsa sopiviksi. Toisaalta tutkimuksen voisi suorittaa myös toisella kieliparilla: onko Suomen säädöksistä

tehdyissä englanninkielisissä käännöksissä mahdollisesti havaittavissa yksinkertaistumista alun perin englanniksi kirjoitettuihin teksteihin verrattuna? Entä päteekö yksinkertaistumishypoteesi Suomen ruotsinkielisiin säädöksiin, jos niitä vertaillaan alun perin ruotsiksi kirjoitettujen lakitekstien kanssa?

Näiden lisäksi olisi mahdollista lähteä tutkimaan muita käännösuniversaaliteorian hypoteeseja samantapaisesta aineistosta. Esiintyykö Euroopan unionin suomenkielisissä lakiteksteissä enemmän vai vähemmän suomenkielelle tyypillisiä piirteitä kuin Suomen säädösteiksteissä? Entä ovatko tietyt ilmaisut tai sanat yliedustettuina? Myös käännöksissä käytetyn kielen kehittyminen vuosien aikana olisi mielenkiintoinen tutkimuskohde. Käännettiinkö vuonna 1995 eri tavalla kuin vuonna 2015? Löytyykö näiden tekstien välillä eroja leksikaalisessa tiheydessä tai niiden sana-sane -suhteissa? Entä onko suomeksi kirjoitetuissa teksteissä tapahtunut samaan aikaan kielellisiä muutoksia, jotka saattaisivat olla merkki Euroopan unionille tyypillisten kielikonventioiden leviämisestä suomenkieliseen lainsäädäntöön?

Myös Euroopan unionin käännösten välinen suhde voisi olla mielenkiintoinen tutkimuskohde. Esimerkiksi Kaisa Koskinen mainitsee, että yksi viimeaikaisista, käännöstyön kannalta positiivisista kehityksistä on ollut se, ettei lainsäädännön käännöstopaa enää automaattisesti sovelleta muita Euroopan unionin tekstejä käännettäessä (2006, 11). Tätä olisi mielenkiintoista tutkia konkreettisesti vertailemalla lakikäännöksiä ja esimerkiksi tiedotteiden käännöksiä sekä niiden leksikaalisia tiheyksiä ja sana-sane -suhteita.

Kvalitatiivisen tutkimuksen puolella mahdollisuuksia on myös monia. Euroopan unionin käännöstoimintaa ja sen kääntäjiä on tarkasteltu ennenkin (esim. Koskinen 2008), mutta kääntäjien käyttämien käännösstrategioiden tarkempi tutkimus voisi olla saatujen tulosten valossa mielenkiintoista, joskin tätä voitaisiin tutkia myös kvantitatiivisesti rinnakkaiskorpuksia käyttämällä. Myös käännösmuisteja ja muiden tietokoneavusteisten käännöstyökalujen käyttöä koskeva tutkimus olisi mielenkiintoinen tutkimuskohde tästä tutkimuksesta saatujen tuloksien nojalla. Se, miten kääntäjät käännösmuisteja käyttävät sekä se, miten niitä ylläpidetään ja editoidaan, voisivat olla hyvin mielenkiintoisia ja havainnollistavia tutkimuskohteita. Lisäksi olisi mielenkiintoista tutkia lakilingvistien roolia kääntämisprosessissa sekä sitä, minkälaisessa yhteydessä he ovat Euroopan unionissa kääntävien kääntäjien sekä muiden suomalaisten asiantuntijoiden, kuten valtion virkailijoiden kanssa. Myös kääntäjien kohdalla olisi mielenkiintoista selvittää ammattikunnan

ylittäviä suhteita: käyttävätkö he esimerkiksi suomenkielisiä lakitekstejä rinnakkaisteksteinä kääntäessään Euroopan unionin lakitekstejä?

Jatkotutkimusmahdollisuuksia on siis runsaasti, ja vaikka tutkimuksen laajuus onkin ollut jokseenkin pieni, on se nostanut esille uusia potentiaalisia tutkimuskohteita, joiden tutkiminen eri metodein voidaan hankkia enemmän tietoa niin käännösuniversaaleista, lain sekä hallinnon teksteistä kuin instituutioista käännöskulttuurina.

LÄHTEET

1.1. KIRJALLISUUSLÄHTEET

- Baker, Mona 1993. Corpus Linguistics and Translation Studies – Implications and Applications. Baker, Mona ym. (toim.), *Text and Technology: In Honour of John Sinclair*. John Benjamins Publishing Company Philadelphia/Amsterdam. 233–250.
- Baker, Mona 1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target* 7:2, 1995. John Benjamins B.V., Amsterdam. 223–243.
- Bernardini, Silvia & Zanettin, Federico 2004. When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translation universals. Mauranen, Anna & Kujamäki, Pekka (toim.). *Translation Universals: Do they exist?* John Benjamins Publishing Company Amsterdam/Philadelphia, 51–62.
- Biber, Douglas, Conrad, Susan & Reppen, Randi 1998. *Corpus linguistics: investigating language structure and use*. Cambridge University Press, Cambridge. 4. painos.
- Cambridge Dictionary of Sociology 2006. Toim. Bryan S. Turner. Cambridge University Press, Iso-Britannia.
- Cao, Deborah 2007. *Translating Law*. Multilingual Matters LTD, Lontoo.
- Chesterman, Andrew 2004. Beyond the particular. Mauranen, Anna & Kujamäki, Pekka (toim.). *Translation Universals: Do they exist?* John Benjamins Publishing Company Amsterdam/Philadelphia. 33–49.
- Chesterman, Andrew 2010. Why study translation universals? *Acta Translatologica Helsingiensia (ATH)*, Vol 1, 38–48.
- Eskola, Sari 2002. *Syntetisoivat rakenteet käännössuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä*. Joensuun yliopiston humanistisia julkaisuja, n:o 30. Joensuun yliopistopaino, Joensuu.
- Eskola, Sari 2004. Untypical frequencies in translated languages. A corpus-based study on a literary corpus of translated and non-translated Finnish. Mauranen, Anna & Kujamäki, Pekka (toim.). *Translation Universals: Do they exist?* John Benjamins Publishing Company Amsterdam/Philadelphia, 83–99.
- Eskola, Sari 2005. Lauserakenteiden käytön piirteitä suomennetussa kaunokirjallisuudessa. Mauranen, Anna & Jantunen, Jarmo H. (toim.) *Käännössuomeksi. Tutkimuksia suomennosten kielestä*. Tampere University Press, Tampere. 225–243.
- Even-Zohar, Itamar 1987. Polysystem Theory. *Poetics Today*, 1:1–2, 287–310.
- Even-Zohar, Itamar 1990. The Position of Translated Literature within the Literary Polysystem. *Poetics Today*, 11:1 1990, 45–51.
- Euroopan unionin perusoikeuskirja. Euroopan unionin virallinen lehti, n:o C 364, 18.12.2000.
- Euroopan unionista tehdyn sopimuksen konsolidoitu toisinto, Euroopan unionin virallinen lehti, n:o C 326, 26.10.2012. 1–390.
- Hallintolaki 6.6.2003/434

- Heikkinen, Vesa & Hiidenmaa, Pirjo & Tiililä, Ulla 2000. *Teksti työnä, virka kielenä*. Gaudeamus, Helsinki.
- Heikkinen, Vesa, Lehtinen, Outi & Mikko Lounela 2001. Kuvia kirjoitetusta suomesta. *Kielikello* 3/2001, 12–15.
- Hermans, Theo 1997. Translation as institution. Snell-Hornby, Mary ym. (toim.), *Translation as intercultural communication*. John Benjamin Publishing Company, Philadelphia. 3–20.
- Hermans, Theo 2012. Chapelle, Carol A. (toim.), *The encyclopedia of applied linguistics*. Wiley-Blackwell, New York. 1681–1686.
- Hodgson, Geoffrey M. 2006. What are institutions? *Journal of Economic Issues* 40:1. 1–25.
- Holmes, James S. 2005 (1972). The Name and Nature of Translation Studies. van den Broeck, Raymond (toim.) *Translated! Papers on Literary Translation and Translation Studies*. Rodopi V.B., Amsterdam, Alankomaat
- Jantunen, Jarmo Harri & Eskola, Sari 2002. Käännössuomi kielivarianttina: syntaktisia ja leksikaalisia erityispiirteitä. *Virittäjä* 2/2002, 184–207.
- Jantunen, Jarmo Harri 2002. Käännössuomen leksikaalisesta (epä)tyypillisyydestä. Korpustutkimuksen kertomaa. *Erikoiskielet ja käännösteoria. VAKKI:n julkaisut*, 29, Vaasa. 131–140.
- Jantunen, Jarmo Harri 2004. *Synonymia ja käännössuomi: korpusnäkökulma sananmerkityksisyyden ja kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin*. Joensuun yliopiston humanistisia julkaisuja, n:o 34. Joensuun yliopistopaino, Joensuu.
- Koivusalo, Esko & Huovinen-Nyberg, Liisa 1985. *Selkeä virkakieli*. 4. painos. Valtion koulutuskeskus, Helsinki.
- Koskela, Merja 2007. Tekstilajit tietoverkossa – jotain uutta, jotain vanhaa ja jotain lainattua? *Käännösteoria, ammattikielet ja monikielisyys. VAKKI:n julkaisut*, N:o 35. 134–145.
- Koskinen, Kaisa 2008. *Translating Institutions: An Ethnographic Study of EU Translation*. St. Jerome Publishing, Manchester.
- Koskinen, Kaisa 2011. Institutional Translation. Gambier, Yves & van Doorslaer (toim.), *Handbook of Translation Studies, Volume 2*. John Benjamin Publishing Company, Philadelphia. 54–60.
- Koskinen, Kaisa 2014. Institutional translation: the art of government by translation. *Perspectives: Studies in Translatology*. 22:4. 479–492.
- Kujamäki, Pekka 2013. Käännöskulttuurit: Näkökulma Tampereen translationaalsiin tiloihin. Koskinen, Kaisa (toim.), *Tulkattu Tampere*. Juvenes Print, Tampere. 30–47.
- Laviosa, Sara 1998. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta* 43 4/1998. 557–570.
- Mauranen, Anna & Jantunen, Jarmo H. (toim.) 2005. *Käännössuomeksi. Tutkimuksia suomennosten kielestä*. Tampere University Press, Tampere.
- Mauranen, Anna 2000. Strange strings in translated language. A study on corpora. Olohan, Maeve (toim.) *Intercultural faultlines. Research models in translation studies I. Textual and cognitive aspects*. St. Jerome Publishing, Manchester, Iso-Britannia. 119–141.

- Mauranen, Anna 2004. Corpora, universals and interference. Mauranen, Anna & Kujamäki, Pekka (toim.). *Translation Universals: Do they exist?* John Benjamins Publishing Company Amsterdam/Philadelphia, 65–82.
- Mossop, Brian 1988. Translating institutions: a missing factor in translation theory. *TTR* 1:2. 65–71.
- Nevalainen, Sampo 2004. Köyhtyykö kieli käännettäessä? Mitä taajuuslistat kertovat suomennosten sanastosta. Mauranen, Anna & Jantunen, Jarmo H. (toim.) *Käännössuomeksi. Tutkimuksia suomennosten kielestä*. Tampere University Press, Tampere. 139–160.
- Olohan, Maeve 2004. *Introducing Corpora in Translation Studies*. Routledge, Oxfordshire, Lontoo.
- Paloposki, Outi 2005. Käännössuomen synty. Mauranen, Anna & Jantunen, Jarmo H. (toim.) *Käännössuomeksi. Tutkimuksia suomennosten kielestä*. Tampere University Press, Tampere. 15–31.
- Piehl, Aino 2003: Selkeyttä vain – EU-kieleen. *Kielikello* 4/2003, 19–22.
- Probirskaja, Svetlana 2009. *Rajankäyntiä. Suomen ja Venäjän kahdenväliset valtiosopimukset käännöstieteellisen avainsana-analyysin valossa*. Acta Universitatis Tamperensis; 1389, Tampere University Press, Tampere.
- Puurtinen, Tiina 1995. *Linguistic acceptability in translated children's literature*. Joensuun yliopiston humanistisia julkaisuja, n:o 15. Joensuun yliopistopaino, Joensuu.
- Robinson, William 2012. *Drafting European Union Legislation*. Euroopan Parlamentti, Bryssel, Saksa.
- Robinson, William 2014. Evolution of European Union Legislative Drafting. *The Loophole*, 7–28.
- Ruonala, Marko & Heli Pietilä 2014. *EU-perusteos*. Ulkoasianministeriön Eurooppatiedotus. Kirjapaino Grano Oy, Kuopio.
- Suomen kielilaki 6.6.2003/423
- Suomen perustuslaki 11.6.1999/731
- Tiilikä, Ulla 2001. Näkymiä tekstityöhön ja työteksteihin. *Kielikello* 3/2001, 12–15.
- Tirkkonen-Condit, Sonja 2004. Unique items – over- or under-represented in translated language? Anna & Kujamäki, Pekka (toim.). *Translation Universals: Do they exist?* John Benjamins Publishing Company Amsterdam/Philadelphia, 177–184.
- Tirkkonen-Condit, Sonja 2005. Häviävätkö uniikkiaineekset käännössuomesta? Mauranen, Anna & Jantunen, Jarmo H. (toim.) *Käännössuomeksi. Tutkimuksia suomennosten kielestä*. Tampere University Press, Tampere. 123–137.
- Toury, Gideon 1995. Prescriptive Translation Studies – and beyond. John Benjamins Publishing Company Amsterdam/Philadelphia.
- Tymoczko, Maria 1998. Computerized Corpora and the Future of Translation Studies. *Meta*, 4/1998, 652–660.
- Vehmas-Lehto, Inkeri 1989. Quasi-correctness: a critical study of Finnish translations of Russian journalistic texts. Neuvostoliittoinstituutti, Helsinki.
- Zanettin, Frederico 2011. Translation and corpus design. *SYNAPS – A Journal of Professional Communication*. 26/2011, 14–23.

1.2. MUUT LÄHTEET

- Eduskunta. Lakiehdotuksella on aina kaksi käsittelyä. Viitattu 8.1.2015. Saatavilla osoitteesta: <http://web.eduskunta.fi/Resource.phx/eduskunta/eduskuntatyto/lainsaadanto/esittely.htx>.
- Euroopan komissio 2012. How to write clearly. Viitattu 15.2.2015. Saatavilla pdf-muodossa osoitteesta: http://ec.europa.eu/translation/writing/clear_writing/how_to_write_clearly_en.pdf.
- Euroopan komissio ja DGT 2011. English Style Guide. Viitattu 15.2.2015. Saatavilla pdf-muodossa osoitteesta: http://ec.europa.eu/translation/english/guidelines/documents/styleguide_english_dgt_en.pdf.
- Europa: EU:n virallinen verkkosivusto. Asetukset, direktiivit ja muut säädökset. Saatavilla osoitteesta: http://europa.eu/eu-law/decision-making/legal-acts/index_fi.htm.
- Europarl: Euroopan parlamentin virallinen verkkosivusto. Tavallinen lainsäätämismenettely. Viitattu 16.1.2015. Saatavilla pdf-muodossa osoitteesta: http://www.europarl.europa.eu/external/appendix/legislativeprocedure/europarl_ordinarylegislativeprocedure_complete_text_fi.pdf.
- Joint Practical Guide for persons involved in the drafting of European Union legislation 2013. Euroopan parlamentin, neuvoston sekä komission lakipalvelut. 2. painos. Viitattu 16.2.2015. Saatavilla pdf-muodossa osoitteesta: <http://eur-lex.europa.eu/content/pdf/techleg/joint-practical-guide-2013-en.pdf>.
- Lainkirjoittajan opas 2013. Oikeusministeriö. Viitattu 9.1.2015. Saatavilla osoitteesta: http://www.oikeusministerio.fi/fi/index/julkaisut/julkaisuarkisto/1378370560107/Files/Lainkirjoittajan_opas_low_20130904.pdf.
- Monikielisyyden puitestrategia 2005. Euroopan unionin komissio. Viitattu 15.2.2015. Saatavilla osoitteesta: <http://eur-lex.europa.eu/legal-content/FI/TXT/?uri=CELEX:52005DC0596>.
- Oikeusministeriö. Lainvalmistelun prosessiopas. Finlex, Edita Publishing Oy. Viitattu 9.1.2015. Saatavilla osoitteesta: <http://lainvalmistelu.finlex.fi/?print>.
- Pulla, Johanna 2011. *Käännösuniversaalit ja temporaalirakenne talouskielessä*. Pro gradu -tutkielma. Tampereen yliopisto.
- Reiman, Mikael 2014. EU ja kääntäminen. *Institutionaaliset tekstit ja niiden kääntäminen*, luentosarja Tampereen yliopistossa, 28.10.2014.
- Schmid, Helmut. TreeTagger - a language independent part-of-speech tagger. Ludwig Maximilian University, Munich, Saksa.
- Scott, Mike, 2012. WordSmith Tools versio 6, Liverpool: Lexical Analysis Software.
- Wagner, Emma 2010. Why does the Commission need a Clear Writing campaign? *Languages and translation*. Directorate-General for Translation 1/2010, 4–5.

LIITTEET

LIITE A

Kaikki tekstit ovat julkisesti saatavilla Edita Publishing oy:n ylläpitämästä FinLex -tietokannasta osoitteesta <<http://www.finlex.fi/>>.

Elatustukilaki 29.8.2008/580

Hallintolaki 6.6.2003/434

Henkilötietolaki 22.4.1999/523

Jäteverolaki 17.12.2010/1126

Kuluttajaturvallisuuslaki 22.7.2011/920

Laki elinkeino-, liikenne- ja ympäristökeskuksista 20.11.2009/897

Laki henkilötietojen käsittelystä rangaistusten täytäntöönpanossa 24.5.2002/422

Laki ihmisen elimien, kudoksien ja solujen lääketieteellisestä käytöstä 2.2.2001/101

Laki korkeimmasta oikeudesta 26.8.2005/665

Laki kuluttajaneuvonnasta 5.12.2008/800

Laki liikepankeista ja muista osakeyhtiömuotoisista luottolaitoksista 28.12.2001/1501

Laki makeisten, jäätelön ja virvoitusjuomien valmisteverosta 17.12.2010/1127

Laki oikeudenkäynnistä rikosasioissa 11.7.1997/689

Laki palvelujen tarjoamisesta 22.12.2009/1166

Laki sosiaalihuollon asiakkaan asemasta ja oikeuksista 22.9.2000/812

Laki sähköisestä lääkemääräyksestä 2.2.2007/61

Laki talous- ja velkaneuvonnasta 4.8.2000/713

Laki terveydenhuollon laitteista ja tarvikkeista 24.6.2010/629

Laki tilaajan selvitysvelvollisuudesta ja vastuusta ulkopuolista työvoimaa käytettäessä 22.12.2006/1233

Laki työsuojelun valvonnasta ja työpaikan työsuojeluyhteistoiminnasta 20.1.2006/44

Laki uusiutuvilla energialähteillä tuotetun sähkön tuotantotuesta 30.12.2010/1396

Laki vaarallisten kemikaalien ja räjähteiden käsittelyn turvallisuudesta 3.6.2005/390

Laki valtakunnanoikeudesta ja ministerivastuuasioiden käsittelystä 25.2.2000/196

Laki viranomaisten tietojärjestelmien ja tietoliikennejärjestelyjen tietoturvallisuuden arvioinnista
22.12.2011/1406

Luonnonsuojelulaki 20.12.1996/1096

Maksupalvelulaki 30.4.2010/290

Merenkulun ympäristönsuojelulaki 29.12.2009/1672

Passilaki 21.7.2006/671

Ryhmäkannelaki 13.4.2007/444

Säästöpankkilaki 28.12.2001/1502

Työaikalaki 9.8.1996/605

Työsopimuslaki 26.1.2001/55

Työterveyshuoltolaki 21.12.2001/1383

Työturvallisuuslaki 23.8.2002/738

Vakuutusyhtiölaki 18.7.2008/521

Vesihuoltolaki 9.2.2001/119

LIITE B

Kaikki tekstit ovat julkisesti saatavilla EUR-Lex -tietokannasta osoitteesta <http://eur-lex.europa.eu/>.

Euroopan parlamentin ja neuvoston asetus (EU) N:o 1093/2010, annettu 24 päivänä marraskuuta 2010, Euroopan valvontaviranomaisen (Euroopan pankkiviranomainen) perustamisesta sekä päätöksen N:o 716/2009/EY muuttamisesta ja komission päätöksen 2009/78/EY kumoamisesta

Euroopan parlamentin ja neuvoston asetus (EU) N:o 1215/2012, annettu 12 päivänä joulukuuta 2012, tuomioistuimen toimivallasta sekä tuomioiden tunnustamisesta ja täytäntöönpanosta siviili- ja kauppaoikeuden alalla

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EU) N:o 181/2011, annettu 16 päivänä helmikuuta 2011, matkustajien oikeuksista linja-autoliikenteessä sekä asetuksen (EY) N:o 2006/2004 muuttamisesta

Euroopan parlamentin ja neuvoston asetus (EU) N:o 260/2012, annettu 14 päivänä maaliskuuta 2012, euromääräisiä tilisiirtoja ja suoraveloituksia koskevista teknisistä ja liiketoimintaa koskevista vaatimuksista sekä asetuksen (EY) N:o 924/2009 muuttamisesta

Euroopan parlamentin ja neuvoston asetus (EU) N:o 492/2011, annettu 5 päivänä huhtikuuta 2011, työntekijöiden vapaasta liikkuvuudesta unionin alueella

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EU) N:o 524/2013, annettu 21 päivänä toukokuuta 2013, kuluttajariitojen verkkovälitteisestä riidanratkaisusta sekä asetuksen (EY) N:o 2006/2004 ja direktiivin 2009/22/EY muuttamisesta (verkkovälitteistä kuluttajariitojen ratkaisua koskeva asetus)

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EU) N:o 648/2012, annettu 4 päivänä heinäkuuta 2012, OTC-johdannaisista, keskusvastapuolista ja kauppätietorekistereistä

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EU) N:o 655/2014, annettu 15 päivänä toukokuuta 2014, eurooppalaisen tilivarojen turvaamismääräysmenettelyn käyttöön ottamisesta rajat ylittävää velkojen perintää varten siviili- ja kauppaoikeudellisissa asioissa

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EU) N:o 910/2014, annettu 23 päivänä heinäkuuta 2014, sähköisestä tunnistamisesta ja sähköisiin transaktioihin liittyvistä luottamuspalveluista sisämarkkinoilla ja direktiivin 1999/93/EY kumoamisesta

Euroopan parlamentin ja neuvoston asetus (EY) N:o 1013/2006, annettu 14 päivänä kesäkuuta 2006, jätteiden siirrosta

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EY) N:o 2006/2004, annettu 27 päivänä lokakuuta 2004, kuluttajansuojalainsäädännön täytäntöönpanosta vastaavien kansallisten viranomaisten yhteistyöstä ("asetus kuluttajansuojaa koskevasta yhteistyöstä")

Euroopan parlamentin ja neuvoston asetus (EY) N:o 45/2001, annettu 18 päivänä joulukuuta 2000, yksilöiden suojelusta yhteisöjen toimielinten ja elinten suorittamassa henkilötietojen käsittelyssä ja näiden tietojen vapaasta liikkuvuudesta

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EY) N:o 593/2008, annettu 17 päivänä kesäkuuta 2008, sopimusvelvoitteisiin sovellettavasta laista (Rooma I)

Euroopan parlamentin ja neuvoston asetus (EY) N:o 66/2010, annettu 25 päivänä marraskuuta 2009, EU-ympäristömerkistä

Euroopan parlamentin ja neuvoston asetus (EY) N:o 726/2004 annettu 31 päivänä maaliskuuta 2004, ihmisille ja eläimille tarkoitettuja lääkkeitä koskevista yhteisön lupa- ja valvontamenettelyistä ja Euroopan lääkeviraston perustamisesta

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EY) N:o 765/2008, annettu 9 päivänä heinäkuuta 2008, tuotteiden kaupan pitämiseen liittyvää akkreditointia ja markkinavalvontaa koskevista vaatimuksista ja neuvoston asetuksen (ETY) N:o 339/93 kumoamisesta

Euroopan parlamentin ja neuvoston asetus (EY) N:o 767/2008, annettu 9 päivänä heinäkuuta 2008, viisumitietojärjestelmästä (VIS) ja lyhytaikaista oleskelua varten myönnettäviä viisumeja koskevasta jäsenvaltioiden välisestä tietojenvaihdosta (VIS-asetus)

EUROOPAN PARLAMENTIN JA NEUVOSTON ASETUS (EY) N:o 864/2007, annettu 11 päivänä heinäkuuta 2007, sopimukseen perustumattomiin velvoitteisiin sovellettavasta laista (Rooma II)

KOMISSIION ASETUS (EY) N:o 1569/2007, annettu 21 päivänä joulukuuta 2007, menetelmästä kolmansien maiden arvopapereiden liikkeeseenlaskijoiden soveltamien tilinpäätösstandardien vastaavuuden määrittämiseksi Euroopan parlamentin ja neuvoston direktiivien 2003/71/EY ja 2004/109/EY nojalla

Neuvoston asetus (EY) N:o 1206/2001, annettu 28 päivänä toukokuuta 2001, jäsenvaltioiden tuomioistuinten välisestä yhteistyöstä siviili- ja kauppaoikeudellisissa asioissa tapahtuvassa todisteiden vastaanottamisessa

Neuvoston asetus (EY) N:o 2201/2003, annettu 27 päivänä marraskuuta 2003 tuomioistuimen toimivallasta sekä tuomioiden tunnustamisesta ja täytäntöönpanosta avioliittoa ja vanhempainvastuuta koskevissa asioissa ja asetuksen (EY) N:o 1347/2000 kumoamisesta

NEUVOSTON ASETUS (EY) N:o 2252/2004, annettu 13 päivänä joulukuuta 2004, jäsenvaltioiden myöntämien passien ja matkustusasiakirjojen turvatekijöitä ja biometriikkaa koskevista vaatimuksista

Neuvoston asetus (EY) N:o 338/97, annettu 9 päivänä joulukuuta 1996, luonnonvaraisten eläinten ja kasvien suojelusta niiden kauppaa sääntelemällä

NEUVOSTON ASETUS (EY) N:o 4/2009, annettu 18 päivänä joulukuuta 2008, toimivallasta, sovellettavasta laista, päätösten tunnustamisesta ja täytäntöönpanosta sekä yhteistyöstä elatusvelvoitteita koskevissa asioissa

ENGLISH SUMMARY

Introduction

In the 1990's, the translation universal hypothesis became a very popular topic of research. At the same time, in 1995, Finland joined the European Union, which meant that the importance of institutional translation and legal translation in Finnish culture and law-making grew. However, although a large body of research exist on both translation universals and institutional texts and institutional translation, for some reason, there is not much research on the topic of translation universals in institutional translations.

This is why, in this study, I intend to study the translation universal hypothesis, and especially the simplification hypothesis, from the point of view of institutional texts. I will conduct my research by utilising corpus tools and two comparative corpora, one of which comprises of Finnish acts, whereas the other corpus includes a selection of Finnish language regulations from the European Union. Based on the previous research, the hypothesis is that some indication pointing to linguistic simplification of the text, such as a heightened type-token ratio and lower lexical density, is found in the translation corpus.

Institutions and their texts

Institution is a difficult concept to define. It can refer to multiple different things, such as social institutions like church and family, to more concrete institutions such as state (The Cambridge Dictionary of Sociology 2006, 300), and despite the fact that its use in social sciences can be traced back to the 18th century, no generally accepted definition exists (Hodgson 2006, 1). This is true in the field of translation studies as well, and many different viewpoints exist. For example, according to Theo Hermans (1997), translation in and of itself can be defined as a cultural institution, since in every culture, one can recognise a number of concrete agents and conventions (1997, 5), which is one possible way to define a translation institution. Brian Mossop has similar thoughts: according to him, most translation activity happens inside some institution, since social structures, even those which work outside established governmental structures, can be classified as institutions in the broad sense

– such as publishing houses (Mossop 1988, 69–70). Thus, it is clear that the notion of institutional translation can be seen as very wide-ranging concept.

However, for the purposes of translation which happens in concrete, governing institutions such as the European Union, it is more feasible to define institutional translation in this more specialised context. From this point of view, it is more convenient to define translation in relation to the institution itself – which is what Kaisa Koskinen (2008) does in her definition of institutional translation. According to Koskinen, institutional translation is an act of communication where the institution communicates with another party in its official capacity (Koskinen 2008, 22). By her definition, institutional translation is the way for the institution to be heard, and during this process, the institution itself gets translated (*ibid.*).

The texts of (and by) public administration – or, in other words, institutional texts – are texts which dictate where and how disputes are settled and the decisions made in the society, and it is used to carry out public administrative duties (Hiidenmaa 2000, 35). This is often why the documents provided by different governing bodies sound official and impersonal (Koivusalo and Huovinen-Nyberg 1985, 11). In addition to the specific style of writing, some subsets of institutional texts, such as law texts, have specific vocabularies and particular expressions, which are specific to the legal system and culture they have been written in (Koivusalo and Huovinen-Nyberg 1985, 16–17). Moreover, the language of law and administration is performative by its nature: words such as “The defendant is found guilty” do not carry just a communicative function, but the abovementioned speech action is in and of itself a concrete action with significant consequences (Cao 2007, 23; Hiidenmaa 2000, 35).

The language of administration and law also has concrete linguistic characteristics: the English language of the uses archaic words (Cao 2007, 21), Finnish language of law uses repetition (Oikeusministeriö 2013, 442), and both are characterised by word strings and long, complicated sentences. Furthermore, according to a part-of-speech analysis conducted by Pirjo Hiidenmaa, one of the characteristics of Finnish administrative texts is the high frequency of nouns, and a high level of lexical density – i.e. the relation of grammatical words to the total number of words. This is to say that they have a high amount of information packed into one sentence. (Hiidenmaa 2000, 44–45).

Institutional texts also have similar textual properties in the way they are connected to other administrative texts: laws are based on and refer to other laws; and the law texts are usually based on reports, analyses, statements from interest groups and so on (Hiidenmaa 2000, 39; Koskela 2007, 134). Merja Koskela calls these relations between different text genres and their relationships with each other a *text genre chain* (FI. *tekstilajiketju*) (2007, 134); the texts the organisation or institution produces are inherently interconnected, which in turn, according to Hiidenmaa, makes administrative texts harder to understand – to understand one text one should understand the texts that have preceded it and be able to perceive its role as a part of the whole process (2000, 39).

All the above mentioned characteristics are reasons why the field of institutional translation and legal translation is perceived as difficult: the systemic differences between specific languages and their linguistic structures, the specialised terminology of law, and its relation to the juridical culture and history, make the task difficult (Cao 2007, 23–25). On the other hand, the concept of equivalence is often emphasised in institutional translations, because when institutions translate, it is often important for them to maintain that the different versions are all equally valid (Koskinen 2011, 57–58), which, naturally, limits the translation strategies available to the translator. These issues are present in Finland and the European Union both: in Finland, legislative texts are translated to Swedish, the second official language of Finland, and in the European Union, the respect of Europe's linguistic diversity has been written into the Treaty on European Union and the Treaty establishing the European Community (article 2). Both institutions have similarities in how they draft legislation, and how they translate it into other languages.

In Finland, most legislative work is done in commissions and working groups, where reports, drafts and other documents are written and stakeholders are heard before the draft is given to the cabinet and the parliament for discussion and vote (Oikeusministeriö). The legislative process in the European Union is very similar: the initial draft is done by policy officers and technical experts, who carry out external consultations during the initial drafting process (Robertson 2014, 11), after which the negotiations and work on the proposal continue in the European Parliament's committees and in the Council (Europarl). They also have similar policies regarding translations, and in both institutions, translations are produced along the legislative drafting process: in the European Union the texts are translated into all official languages before they are put forward to the Parliament and the Council, whereas in Finland the draft is translated into Swedish before the draft is given to the cabinet. However, in both cases, the translations only come in after one "complete" version has already been

written, and neither institution uses multilingual drafting process, where different language versions are worked on simultaneously. This is to say that, on the one hand, the processes are so similar that it is logical to propose that these similarities in the drafting process are reflected in the texts themselves. On the other hand, however, the process of translating these texts is closer to the traditional legal translation than multilingual legislative drafting, so from this point of view it could be assumed that they have some linguistic qualities which are often assigned to translated texts. This dichotomy is why I personally find the topic of legal translation an intriguing point of discussion from the point of view of descriptive translation studies in general, and from the point of view of the translation universal hypothesis in particular.

Translation universals

The translation universal hypotheses have their roots in descriptive translation studies, which, in turn, has been influenced by Itamar Even-Zohar's polysystem theory. According to Even-Zohar's polysystem theory, semiotic phenomena, such as language and literature, should be seen as interconnected, heterogeneous systems, instead of separate elements (1979, 287). He also asks if it is possible that translations form a similar interconnected cultural and linguistic "network of relations" that original literature is hypothesised to form (1990, 45–46). This tradition of looking at translations from the systemic-cultural perspective is continued by Toury (1995), who argues that a translator should be seen as having a social role in his or her culture, and as operating according to the norms of that culture (Toury 1995, 53). According to him, the act of translation includes at least two languages and cultures, which have their own norms whose influence, in turn, can be seen in the translations. This is why, according to Toury, we can learn about the different cultural norms translations are expected to adhere to in any given culture by studying these translations themselves (Toury 1995, 56).

From this point of view he divides translation norms into different categories: to *preliminary norms*, which govern what is translated (i.e. translation policy), and to *operational norms*, which govern the process of translation itself. He then divides operational norms further into *matricial norms*, which govern the amount of material which is intended to substitute the corresponding source-language material (i.e. translations phenomena such as additions and omissions), and to *textual-linguistic norms*, which govern what kind of linguistic material is used in the translation, which is to say, what

kind of linguistic strategies the translator uses in the translation process. Moreover, one way to study these norms, according to Toury, is to study the translations themselves. (Toury 1995, 58–59). In the end, he suggests a hypothesis that there are characteristics which are typical to translations, and forms hypothetical laws of translational behaviour, such as the “*law of growing standardisation*” (Toury 2015, 267) and the “*law of interference*” (ibid. 274– 275).

Toury’s theory and his perspective which regards translation as cultural activity governed by norms and laws was adopted by a number of researches, including Andrew Chesterman, Mona Baker, and Theo Hermans. In 1993, citing Toury’s and Even-Zohar’s works, Baker made a suggestion of using corpora to study these translation norms, and she mentions the universal characteristics of translations as one potential topic of research (Baker 1993, 243). She mentions phenomena such as explicitation, simplification, the tendency to avoid repetition, and the tendency to exaggerate the features of the target language as possible universal characteristics of translation (ibid. 244). Andrew Chesterman, in his 2004 paper, classified the potential translation universals into *source universals* – i.e. potential universal characteristics in which the translations differ from the source text – and into *target universals* – i.e. the potential universal characteristics in which the translations differ from the works originally written in the target language (2004, 37).

The translation universal hypotheses are not without problems, of course. One is a question of definitions: can we truly call anything “universal” if we do not investigate all the possible avenues and languages pairs of translation? And what, exactly, do we consider to be a translation “worthy” enough to be considered for study – in effect, should we include “bad” translations (Chesterman 2004, 42)? Also, according to Chesterman, one issue with the translation universals hypotheses is that oftentimes the terms are not properly defined: researches might, for example, investigate source universals from a monolingual corpus, which is problematic according to Chesterman, since logically one cannot find source universals by studying material which does not include the source text itself (Chesterman 2004, 43).

Nevertheless, the translation universals hypotheses have been studied extensively by, to name a few, Anna Mauranen (2000, 2005), Jarmo Jantunen (2002, 2004), Sonja Tirkkonen-Condit (2004, 2005), Sari Eskola (2002, 2004, 2005), and Sara Laviosa (1998). Tirkkonen-Condit, for example, studied clusters of Finnish words with the suffixes *-kin*, which have no direct equivalent in English and other

Indo-European languages. According to her findings, the *-kin* suffixes were more clustered in translations than in texts which had been originally written in Finnish, which she interpreted to support the proposed translation hypothesis of simplification (Tirkkonen-Condit 2005, 133–135). Sampo Nevalainen, who studied the vocabulary of texts translated into Finnish by using frequency lists, came to the conclusion that the percentage of the most frequently used words in the text was higher in the Finnish translations – and, like Tirkkonen-Condit, he arrived to the conclusion that the results supported the simplification hypothesis (Nevalainen 2005, 146–145).

Not everyone has come to the same conclusion, however. For example, in his doctoral thesis Jarmo Jantunen came into the conclusion that although there was evidence of untypical lexical frequencies and combinations (2004, 215–216), his research did not find conclusive evidence of the simplification hypothesis (ibid., 218). Similarly, Tiina Puurtinen, who studied the explicitation hypothesis through connectives in translated children's literature, found that although some connectives were more common in translated Finnish than others, complete opposite was true in case of other connectives, and thus she deemed her results as neither supporting nor contradicting the explicitation hypothesis (Puurtinen 2004, 174).

All in all, although the research into translation universal hypothesis has been plentiful, there are still points of discussion and contradictory results: evidence which supports the existence of potential translation universals has been found, but there have been some conflicting and inconclusive studies as well. Another important point is that very few studies which focus on administrative texts and their translations or legal translations exist – which is exactly why these translations ought to be studied.

The corpora, the methods, and the texts

Most research into translation universals have been conducted with the use of corpora. One reason for this might be that Mona Baker explicitly mentioned corpora as a way to study this particular phenomenon (Baker 1993; Baker 1995), but most likely this is the case because a corpus is a particularly useful tool for handling large amounts of data and looking for systematic patterns from said data.

A corpus, according to Baker (1995) is a group of texts in an electronic format, which have been collected for a specific purpose (1995, 225). The main goal of a corpus is to represent language or a certain part of language, which in turn defines what kind of texts are included in the corpus and how widely the results of the studies made with this corpus can be generalised (Biber et. al. 1998, 246). This is why the texts which are selected into the corpus should follow a set criterion to make sure that the corpus is a representative sample of the area of language it is meant to analyse (ibid., 225). Such criteria include, for example, the type of language (written vs. spoken language), geographical differences (American English vs. British English), and the language itself (standard language vs. academic language), and so on (Baker 1995, 229). It is, of course, a very problematic task to gather a representative sample of any part of language, since it would require complete knowledge of all possible linguistic and contextual variations texts can have (Biber et. al 1998, 246), which is of course impossible. In addition, another issues, such as the size of the corpus, can also have a significant effect on the results (ibid., 249). However, the advantages of corpus research include the volume of text that can be studied through the corpus in quantities which would not be practical by using qualitative methods (Biber et. al. 1998, 1–4). In addition to this, corpus makes it possible to conduct extensive analytic research on the structure or language through applications such as part-of-speech tagging.

Corpora can be categorised in different ways according to which they have been collected and used. *Parallel corpora*, for example, are corpora which include the source texts and its translations in one or more languages. However, as Maeve Olohan points out, the term is somewhat problematic, since the term could also potentially be used to refer to a bilingual corpus which is made of text which are *parallel*, which is why the term *translation corpus* has been suggested instead (2004, 24). In addition, in translation studies, the parallel corpora are often *aligned*, which means that their linguistic units (such as sentences) are aligned into the parallel units in the other language(s) (Olohan 2004, 26). *Comparable corpora* are monolingual corpora, which have been widely used in the study of translation universals in particular. The most common way to compile and use a comparable corpus is to compile two corpora: one which consists of material originally written in one language, and another which consists of material translated into the same language.

Several ways to use the corpora exist as well. It is possible to study the *frequencies* of words or expressions, or the *concordances* present in the material. The tools of corpus method include, for example, *frequency lists*, which calculate the words in the corpus and list how many times they are

used in the corpus. Another tool related to frequencies is the *type-token ratio*, which is the ratio of singular words and the amount of words in the corpus. Thus, the type-token ratio can be used to calculate the variety of the vocabulary of the text: a low type-token ratio signifies a relatively simple vocabulary, whereas a high type-token ratio is a sign of varied vocabulary, where any particular word is used less frequently. This is one way to study the potential translation universal of simplification. Potential translation universals such as untypical lexical patterns, on the other hand, can be studied by utilising corpus tools such as *key word in context*, where the corpus shows the key word or phrase which is being studied in context. This makes it possible to analyse collocations, for example.

To sum up, corpus is a particularly suitable research method for studying phenomena such as translation universals, because it offers a way to study large amounts of texts and analyse them in terms of common patterns, which is a way to discover potential translation universals. Corpora can answer the question “What general patterns can be found from these texts”, which is why they are such a popular research method among the researchers who study translation universals. Despite this, corpus as a research method is not without its problems, as discussed above, but most of the issues can be addressed and controlled within the framework of the study. Biber et. al. point out that the most important part of corpus research is to take into account how different categories have been identified and what criteria has been used when the corpus has been categorised (Biber et. al. 1998, 248). Also, as mentioned above, it is important to specify the criteria by which the texts have been selected, and to acknowledge the limitations these criteria impose on the generalisability of the results.

When all the above-mentioned points are taken into account, it becomes clear that corpus is a particularly suitable method to my research into the patterns of legal texts and the possible translation universal of simplification found therein. In my research, I am using comparative corpora of legal texts: the first consists of Finnish acts, and the second consist of Finnish translations of European Union regulations. Most of the selected texts are from the 21st century, and there are no texts which have been written and/or translated before the year 1995, when Finland joined the European Union. The acts and regulations cover a variety of subjects from banking to environmental law and consumer protection so that the results will not be affected by one particular sub-genre of legal writing (business law, criminal law etc.), which would mean that it would be possible to pick up on the general patterns in legal texts as a genre. The introductory phrases and annexes have been excluded to make sure the

corpus reflects the general language these texts use, and that formative issues, such as repetitive introductory phrases, do not distort the results.

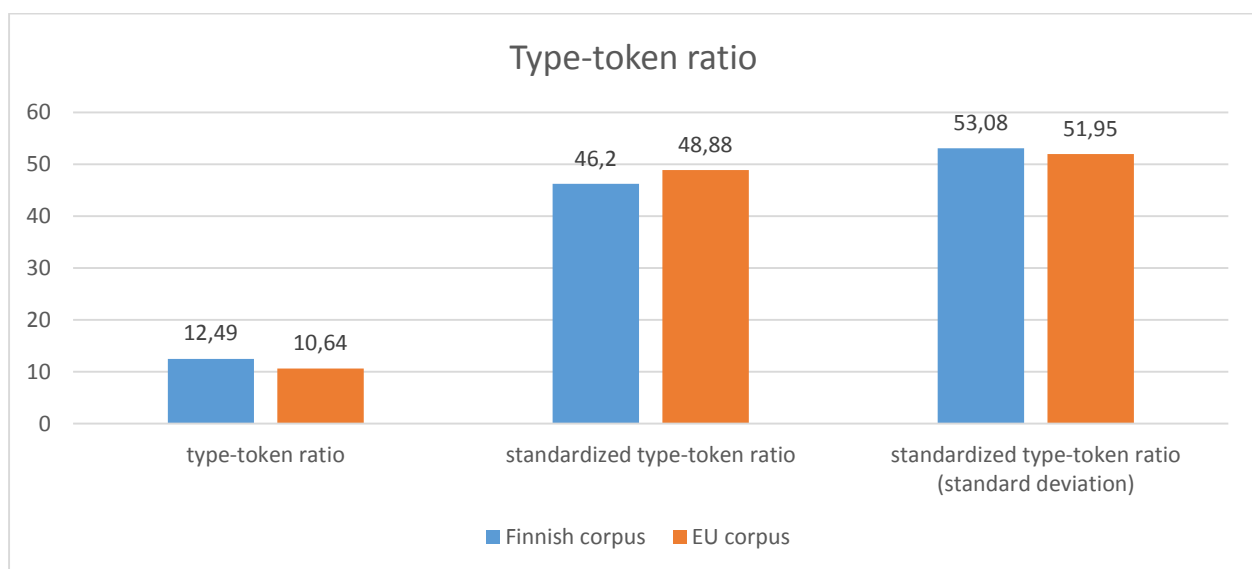
The corpora are somewhat small: the word count in the EU corpus is 202,919 words from 24 unique texts, whereas the Finnish corpus consists of 200,680 words and 36 unique texts. Although the corpora are small, and this has to be taken into account when the results are analysed and the conclusions are generalised on a wider body of texts based on the results, the texts in the corpora have been chosen according to a specific criterion, and they are meant to represent a small part of language, in which case, smaller corpora can be used (Zanettin 2011, 15). The corpora have also been tagged using the TreeTagger programme developed by Helmut Schmid, and the corpus programme used in the analysis is WordSmith Tools, version 6. With these tools, I will be investigating the translation universal hypothesis of simplification through analysing the general lexical properties of the texts, such as sentence and word length, the type-token ratio of the texts, and their lexical density and part-of-speech distribution. Based on previous research, the hypothesis is that the type-token ratio and the lexical density of the translated texts will be lower (see Nekalainen 2005; Laviosa 1998).

Analysis

The analysis shows that the average sentence length is much longer in the EU corpus than in the Finnish corpus: the standard deviation of sentence length measured in words in the EU corpus is 34.60 words, whereas in the Finnish corpus the number is 24.20 words. In addition, the difference in percentages between the average sentence length and the standard deviation is much higher in the EU corpus (7.9 per cent vs. 17.6 per cent) which would indicate not only that the EU translations have longer sentences, but also that the sentence length varies more in the EU texts than it does in the Finnish legal texts. Both texts also have a high number of long words: the average word length in the EU corpus is 8.91 characters, whereas in the Finnish corpus the number is slightly lower 8.20 characters. However, the percentage of long word of more than 9 characters in length is maybe somewhat surprisingly slightly higher in the Finnish corpus (37.5 per cent vs. 36.9 per cent). However, these differences are not significant enough to draw any significant conclusions.

The only significant finding is the longer sentence length in the translated texts, which is an observation Sara Laviosa (1998, 561) and Svetlana Probirskaja (2009, 135) have made as well. This could be a sign of “lengthening”, which Andrew Chesterman categorises as a potential source universal (2004, 40), but, as was mentioned earlier, it would be problematic to identify this phenomena as a translation universal, since it is a source universal and when the corpora are monolingual, they are designed to study target universals. However, another things to consider are the translation guidelines of the European Union legal translation, which state that all the different language versions must have equal amount of sentences (Piehl 2003, 19–20), which could explain these differences in sentence length.

The type-token ratio, on the other hand, is almost identical in both corpora, as seen in picture 1.



Picture 1: Type-token ratios

The results are particularly interesting for two reasons. One, they are so similar that it is safe to say no evidence of simplification hypothesis was discovered, which contradicts the existing research. In addition, the results seem to partially confirm the observations made of the structure of the language of the law: the low type-token ratio could be explained by repetition, which, as mentioned above, is one of the characteristics of legal language. The frequency lists support this observation: in the EU corpus, for example, common words include terms such as *artikla* (EN. article) and *jäsenvaltio* (EN. member state), which are unique to EU legal language. In addition, when the frequency lists of the texts in the corpora are studied separately, a similar pattern emerges, and certain words and word

forms related to the subject of the text are repeated. However, this is to be expected when studying such small corpora. In addition, it must also be mentioned that this repetition is also caused by the various forms of the verb *olla*, and the most common conjunctions such as *ja*, *että*, and so on, which is to be expected in a large amount of text, since these are highly common words that appear repeatedly in all texts. Thus, a more thorough, qualitative analysis would be needed to make any concrete statements on the nature of repetition in these texts.

The distribution of parts-of-speech in both corpora are very similar as well (picture 2, next page). Since part-of-speech tagging always carries a margin of error³, it is safe to say that the differences are not significant enough to make any definitive conclusions. The lexical densities of the corpora are 73.7 per cent for the EU corpora, and a slightly lower 69.95 per cent for the Finnish corpora. This finding contradicts with the previous research on the topic (Laviosa 1998), but it must be mentioned that when the size of the corpora, the accuracy of the tagging and the margin of error are taken into account, it would be problematic to claim that the lexical density of the EU texts is significantly higher.

Conclusion

All in all, no significant evidence of the simplification hypothesis was found in the study. This could be because of the small size of the corpora: it is possible there was simply not enough data to find the truly unique characters and the most significant patterns. But the result could also be an indication that translation universals – if they exist – differ from one text type and genre to another. This could be why the results of this study contradict with the previous research: it is possible the results are not comparable, because the corpora are not comparable. It is possible that in this particular case, the same parameters should not – or cannot – be used to study literature and legal translations. A significant body of evidence on translation universal research on literature translation seems to suggest that the features analysed in this study, i.e. simplification and lower lexical density, can be found from literature translations, which would imply that these particular results are either a result

³ The accuracy of the part-of-speech tagging in these corpora has been calculated by hand from a random sample of 200 tags: it has accuracy of approx. 89.5 per cent, with a margin of error of ± 7.7 percentage points with 95 per cent accuracy.

of the genre of the texts and the translations which are being studied, or the result of an insufficient corpus. This would be why the most obvious future research opportunity would be to recreate the study with larger, manually edited and lemmatised corpora.

EU Corpus		Finnish corpus	
Part of speech	Percentage	Part-of-speech	Percentage
Nouns	45.1	Nouns	44.9
Verbs	11.0	Verbs	10.8
Participles ⁴	7.1	Participles	6,9
Adjectives.	8.2	Adjectives	4.9
Conjunctives	7.0	Conjunctives	8.2
Numerals	5.0	Numerals	7.3
Pronouns	4.6	Pronouns	3.4
Adverbs	2.3	Adverbs	2.3
NON-TWOL	6.8	NON-TWOL	8.2
Other	3.0	Other	3.0

Picture 2: Part-of-speech distribution

However, if the findings of this study are a result of the former rather than the latter, the implications of the study open up new and interesting avenues for future translation research. Are there perhaps previously unknown translation universals which only exist in certain genres or sub-genres of texts and their translations? Or, is it possible that the translation universals do not only rely on norms, but on the *absence* of norms – or at least the absence of formal norms – as well: legal translation is a field which is highly regulated and governed by norms, particularly in institutions such as the European Union, where each language version of the text is legally valid under the European Union law,

⁴ Participles have been given their own category because they can be categorised into either verbs or adjectives, depending on the context.

whereas most norms and laws governing the field of literature translation are not as concrete and more cultural and practical in nature.

One possible interesting new avenue of research would also be the role of translation memories, other computer assisted translation tools and term banks in translation and how their use affects the lexical density and type-token ratio of the texts which are produced. One more potential topic of research could be the general translation culture and practices in multilingual institutions. Some research already exists on this topic (see Koskinen 2008), but in light of these findings, it would be interesting to study the topic further and investigate if the translators use Finnish acts as reference material when making translation decisions, or how much they consult Finnish lawyers and legal experts during the translation process.

All in all, some interesting findings were made in this study, even though the results would indicate that no support for the simplification hypothesis was found. Despite this, the study raised some fascinating questions on the nature of translation universals, administrative texts, and the topic of institutional translation in general.